# 遗传算法的数学基础

张文修　梁　怡　编著

西安交通大学出版社

# Mathematical Foundation of Genetic Algorithms

Zhang Wenxiu
Leung Yee

Xi'an Jiaotong University Press

**图书在版编目(CIP)数据**

遗传算法的数学基础/张文修,梁怡编著 —西安:西安交通大学出版社,2000.5
ISBN 7-5605-1256-9

Ⅰ.遗… Ⅱ.①张… ②梁… Ⅲ.遗传算法-数学基础
Ⅳ.O242.23

中国版本图书馆 CIP 数据核字(2000)第 23337 号

*

西安交通大学出版社出版发行
(西安市兴庆南路 25 号 邮政编码:710049 电话:(029)2668316)
西安电子科技大学印刷厂印装
各地新华书店经销

*

开本:850mm×1 168mm 1/32 印张:7.125 字数:173 千字
2000 年 5 月第 1 版 2001 年 4 月第 2 次印刷
印数:3 001~5 000 定价:18.00 元(精) 12.00 元(平)

---

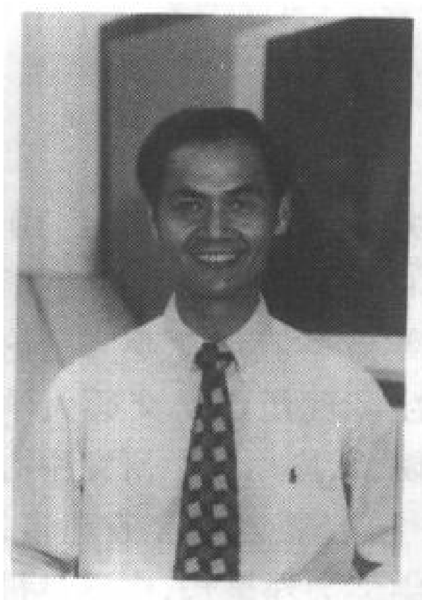
## 作者简介



张文修，教授，1940 年 10 月出生于山西翼城县，1967 年毕业南开大学概率统计专业信息论方向研究生。现任西安交通大学研究生院院长，应用数学研究中心副主任，中国数学学会常务理事，陕西省数学学会理事长，中国模糊集与系统学会副理事长，国际 IFSA 会员，《模糊系统与数学》杂志副主编，《工程数学学报》杂志主编等。

张文修教授长期从事应用概率与集值随机过程、计算机智能推

理基础与计算仿真的理论研究。先后主持和完成5项国家自然科学基金课题和两项国家863高新技术课题。作为项目主持人,张文修获国家级优秀教学成果一等奖一项、二等奖一项、陕西省科技进步一等奖一项、国家教委优秀教材一等奖一项、西安交通大学科技成果一等奖两项。由科学出版社、贵州科技出版社、西安交通大学出版社等先后出版专著和教材12种,在中国科学、数学学报、科学通报、数学进展及国际学术刊物上发表研究论文80多篇。多次出国参加国际学术会议、访问、考察和科研合作。



梁怡,教授,现任香港中文大学地理系讲座教授及系主任,环境研究中心研究员及地球信息科学联合实验室学术委员会副主任。获中文大学社会科学学士及美国科罗来多大学文学硕士,理学硕士及博士学位。担任多种国际杂志编委和学会成员。由国际出版社出版专著2种,西安交通大学出版社出版专著1种。在国际学术刊物和国际会议上发表论文100多篇。研究方向为空间分析、环境工程与专家系统、地理信息系统及决策支持系统、模糊逻辑与神经网络、遗传算法等。先后主持和完成10项RGC(香港政府)课题。已完成的三项课题均获杰出研究奖。被选入以下名人录:(1) Edition Two of 5000 Personalities of the World by the American Biographical Institute; (2) Men and Wom-

en of Distinction (Fourth World Edition) by the International Biographical Centre; (3) Men of Achievement (Fifteenth Edition) by the International Biogreaphical Centre; (4) Dictionary of International Biography (Twenty Second edition) by the International Biographical Centre.

## ABOUT THE AUTHORS

Zhang Wenxiu is a professor of Xi'an Jiaotong University, Xi'an, China. He graduated from Nankai University in 1967, specialized in information theory, probability and statistics major. He currently serves as dean of graduate school and vice-director of Research Centre of Applied Mathematics at Xi'an Jiaotong University, president of Mathematics Society of Shannxi Province, vice-president of Chinese Society of Fuzzy Set and Systems, Vice-editor-in-chief of Chinese journals: "Fuzzy System and Mathematics" and editor-in-chief of "Journal of Engineering Mathematics". He is a member of the Standing Council of Mathematics Society of China, and of the IFSA.

Professor Zhang has been active in research on applied probability theory, set-valued stochastic process and computer reasoning in AI. He is the author or coauthor of over 80 academic journal papers, and 12 textbookds and monographs. As a principal investigator or author, he won prizes of the most distinguished teaching achievement in China in 1993, the excellent textbooks in universities in 1987, the first class award of Scientific Progress in Shannxi Province in 1987 and two first class awards of Research results of Xi'an Jiaotong University in 1988 and 1993. Professor Zhang has held many visting positions in Chinese and foreign universities.

Leung Yee is a professor of Geography and Chairman of the Department of Geography, Research Follow of the Center for Environmental studies, and deputy director of the Academic Committee of the Joint Laboratory for Geoinformation Science of the Chinese University of Hong Kong, He obtained his B. S. Sc. from the Chinese University of Hong Kong in 1992 and M. A., M. S., and PH D from the University of Colorado, USA in 1977.

Dr. Leung has been conducting research on spatial analysis environmental engineering, expert system, Fuzzy programming and neural networks, and evolutionary computation. He has published over 100 journal papers and three monographs "Spatial Analysis and Planning under Imprecision" (Amsterdam, North-Holland, 1988), "Intelligent Spatial Decision Support Systems" (Berlin, Springer-Verlag, 1997), "Principles of Uncentain Reasoning" (Xi'an Jiaotong University, 1996). He is a holder of 10 RGC (Hong Kong SAR) grants and has received "Research Excellence Award" for all of the 3 completed projects. He is a member of the editorial boards of several international journals, and has been listed in the following Who's Who.

## 内容简介

遗传算法(genetic algorithm)是模拟自然界生物进化过程与机制求解问题的一类自组织与自适应的人工智能技术,已广泛应用于计算机科学、人工智能、信息技术及工程实践。

本书重点在于阐述遗传算法的数学基础。全书共分3章,第1章给出了遗传算法的几何理论,第2章给出了遗传算法的马尔可夫链分析,第3章给出了遗传算法的收敛理论。

本书可以作为研究遗传算法的参考书,也可以作为应用数学、计算机科学、系统科学等专业研究生的教材。

# 前　言

遗传算法(genetic algorithm)是模拟自然界生物进化过程与机制求解极值问题的一类自组织、自适应人工智能技术。它模拟达尔文的自然进化论与孟代尔的遗传变异理论,具有坚实的生物学基础;它提供从智能生成过程观点对生物智能的模拟,具有鲜明的认知学意义;它适合于无表达或有表达的任何类函数,具有可实现的并行计算行为;它能解决任何类实际问题,具有广泛的应用价值。因此,最近十多年,遗传算法在国内外倍受重视。

遗传算法的思想由来已久。早在20世纪50年代,一些生物学家就着手于计算机模拟生物的遗传系统。1967年,美国芝加哥大学Holland, J.H.教授在研究适应系统时,进一步涉及进化演算的思考,并于1968年提出模式理论。1975年,Holland教授的专著《自然界和人工系统的适应性》(Adaptation in Nature and Artificial Systems)问世,全面地介绍了遗传算法,为遗传算法奠定了基

础。此后,遗传算法无论在理论研究方面,还是在实际应用方面都有了长足的发展。

一种理论的成熟取决于它的数学描述的完善程度。随着遗传算法的广泛应用,理论研究也越来越深入。对于遗传算法的理论分析基本上有两大类:一类是遗传算法的马氏链模型;一类是Vose-Liepins 模型。表面上这两类方法似乎是有限状态与无限状态之分,但研究的方法大相径庭。对于马氏链模型主要采用转移概率与极限理论;对于 Vose-Liepins 模型主要采用不动点理论,具有浓厚的几何色彩。本书的第 1 章除介绍遗传算法的基本概念以外,介绍了 Vose-Liepins 模型,第 2 章介绍了遗传算法的马氏链模型。不管是遗传算法的马氏链模型,还是 Vose-Liepins 模型,重点都是研究遗传算法的收敛性。我们在第 3 章中,在马氏链模型与Vose-Liepins 模型的基础上,抓住了遗传算法收敛的一些基本条件以及这些基本条件之间的关系,给出了遗传算法收敛的一般定理,建立了证明遗传算法收敛的一般方法。通过对于各种具体遗传算法收敛性的证明,证明了这种方法的有效性。第 3 章的结果主要是西安交通大学理学院信息与系统研究所的专家们与香港中文大学的有关专家合作科研的成果。这些专家包括徐宗本教授、聂赞坎教授、梁广锡教授、高勇副教授、张讲社副教授、段启宏博士等,当然也包括本书的两位作者。读者将可以看出,第 3 章中所提供的方法更加清晰、简明。因此本书除了是对国际上遗传算法理论分析的总结之外,同时也是西安交通大学与香港中文大学两校有关研究成果的总结。感谢这些专家们将尚未发表的结果提供出来,使本书的写作更加丰富。特别要指出的是,本书的绪论主要引自我们指导的博士生高勇的博士论文。

尽管遗传算法已有不少的理论研究,但仍然不能说是很完善,或很深入。理论的研究与遗传算法真实广阔的发展前景之间还有着很大的距离。遗传算法对生物演化的模拟基本上还是形式的,还未深入到生物演化内部规律的模拟,因此这使遗传算法的作用

大受局限。应用与理论的相互推动，将会使遗传算法放射出巨大光芒。当生命演化模拟的遗传算法用来研究生物演化，特别是研究大脑的演化过程时，那将是遗传算法辉煌的时期。致力于遗传算法的研究必须站在科学技术时代的前沿，正因为如此，当完成本书稿时，作者深感本书之不完善——对过去的研究历史挂一漏万，对未来的发展又未能有所展望。不过作为进入研究遗传算法的捷径，本书对那些初涉遗传算法者将会有些方便之处，可以使读者避免重新阅读大量文献，同时给出了较大的思维活动空间。本书如能有这种作用，已倍感荣幸。

作者

1999.5.1

**本专著所从事的研究**

*受国家自然科学基金、国家863(306)基金*
*及国家教育部博士点基金资助*

# 目　　录

# 绪　论

## 0-1　遗传算法是一种仿生优化算法

我们的自然界充满了奇迹，而生命的繁衍生息则是这些奇迹中的奇迹。生命是脆弱的，生命也是顽强的。从远古时代单细胞开始，历经环境变迁的磨难，生命经过了从低级到高级、从简单到复杂的演化之路，不但延续下来，而且产生了人类这样有思维、有智力的高级生命体。人类找到了生命的最佳结构与形式，它不仅仅可以被动地适应环境，更重要的是它能够通过学习、模拟与创造，不断提高自己适应环境的能力。

在人类的历史上，通过学习与模拟来增强自身适应能力的例子不胜枚举。模拟飞禽，人类可以翱游天空；模拟游鱼，人类可以横渡海洋；模拟昆虫，人类可以纵观千里；模拟大脑，人类创造了影响世界发展的计算机。人类的模拟能力并不仅仅局限于自然现象和其他生命体。自从20世纪后半叶以来，人类正在将其模拟的范围延伸向人类自身。神经网络（neural network）是人类对其大脑

信息处理机制的模拟，模糊系统(fuzzy systems)是人类对其思维方式的模拟。除了向自身结构的学习以外，人类还可以向其自身的演化这一更为宏观的过程学习，来增强自己解决问题的能力，其代表性的方法就是遗传算法(genetic algorithm, GA)。

人类之所以能够向其自身的演化学习以增强决策问题的能力，是因为自然演化过程本质就是一个学习与优化的过程。这一优化过程的目的是使生命体达到适应环境的最佳结构与效果。曾经主宰地球的恐龙由于庞大的身躯跟不上环境的变迁而灭绝；长颈鹿为了觅食而长长了脖子；老鼠的机敏是为了生存而挣扎的结果；青蛙的存活则得益于其两栖式左右逢源的能力；人类解放出有力的双手，得益于类人猿求生的努力，而正是这一对不再用于行走的双手，使人类成了这个世界的主宰。自然演化遵循着一种奇妙的规律，这一规律截然不同于人类自己发明的解析方法。我们知道，人体36.8℃的恒定体温是至关重要的，0.5℃的偏差足以使人产生病态。那么在0～100℃这样大的范围内，人体是如何求出如此精确的最佳体温的呢？这绝对不是梯度下降等方法所能算出的，而我们也相信目前的解析方法远远无法完整描述体温、人体结构与环境间错综复杂的关系。自然演化的规律是达尔文发现的，它的核心思想是：物竞天择、适者生存。它对人类的作用与影响不亚于牛顿的万有引力学说，而奇怪的是这一理论时至今日才被应用于工程实践。

遗传算法是一种更为宏观意义下的仿生算法，它模仿的机制是一切生命与智能的产生与进化过程。它通过模拟达尔文"优胜劣汰、适者生存"的原理激励好的结构；通过模拟孟德尔遗传变异理论在迭代过程中保持已有的结构，同时寻找更好的结构。作为一种随机的优化与搜索方法，遗传算法有着其鲜明的特点：

(1) 遗传算法的操作对象是一组可行解，而非单个可行解；搜索轨道有多条，而非单条，因而具有良好的**并行性**。

(2) 遗传算法只需利用目标的取值信息，而无需梯度等高价值信息，因而适用于任何大规模、高度非线性的不连续多峰函数的优化以及无解析表达式的目标函数的优化，具有很强的**通用性**。

(3) 遗传算法择优机制是一种"软"选择，加上其良好的并行性，使它具有良好的**全局优化性**和**稳健性**。

(4) 遗传算法操作的可行解集是经过编码化的(通常采用二进制编码)，目标函数解释为编码化个体(可行解)的适应值，因而具有良好的**可操作性与简单性**。

## 0-2　遗传算法的发展与现状

遗传算法的产生归功于美国 Michigan 大学的 Holland 在 20 世纪 60 年代末、70 年代初的开创性工作，其本意是在人工适应系统中设计的一种基于自然演化原理搜索机制。大约在同一时间，Foegl 和 Rechenberg 及 Schwefel，引入了另两种基于自然演化原理的算法，演化程序(evolutionary programming)和演化策略(evolution strategies)。这三种算法构成了目前演化计算(evolutionary computation)领域的三大分支，它们从不同层次、不同角度模拟自然演化原理，以达到求解问题的目的。

Holland 不仅设计了遗传算法的模拟与操作原理，更重要的是，他运用统计决策理论对遗传算法的搜索机理进行了理论分析，建立了著名的 Schema 定理和隐含并行性(implicit parallelism)原理，为遗传算法的发展奠定了基础。将遗传算法应用于函数优化始于 De Jong，他在其博士论文中设计了一系列遗传算法的执行策略和性能评价指标，对遗传算法性能做了大量的分析。De Jong 的在线(on-line)和离线(off-line)指标仍是目前衡量遗传算法性能的主要手段，而他精心挑选的 5 个试验函数(称作 De Jong's five test functions)也是目前遗传算法数值试验中用得最多的试验函数。

在 Holland 和 De Jong 的工作之后，遗传算法经历了一个相对平稳的发展时期，逐渐被人们所接受和运用。遗传算法的发展高潮开始于 20 世纪 80 年代末，而且延续至今。人们对遗传算法兴趣的日益增长有两个背景：其一是工程领域，特别是人工智能与控制领域，不断涌现出超大规模的非线性系统，在这些系统的研究中存在着大量的经典优化方法所不能有效求解的优化问题，诸如神经网络连接权重及网络拓扑结构的优化、模糊系统中模糊规则的选取及隶属函数的确定、知识库的维护与更新等等；其二，遗传算法本身就是一种模拟自然演化这一学习过程的求解问题方法，它能以独立的或与其它方法相结合的形式用于智能机器学习系统的设计中。经过近 10 年的努力，遗传算法不论是在应用上、算法设计上，还是在基础理论上，均取得了长足的发展，已成为信息科学、计算机科学、运筹学和应用数学等诸多学科所共同关注的热点研究领域。

### 1. 遗传算法在神经网络、模糊系统和机器学习中的应用

在应用中，遗传算法几乎渗透到从工程到社会科学的诸多领域，我们无法在此一一列举。这里简略介绍遗传算法与神经网络、模糊系统和机器学习结合的例子。

神经网络的学习包含了两个优化过程，分别是网络连接权重的优化和网络拓扑结构的优化。优化连接权重最著名的方法是 Rumelhart 提出的基于梯度下降法的反向传播法（backpropagation, BP）。BP 算法的最大弱点是局部极小问题和无法学习网络拓扑结构。作为一种通用性和全局性良好的优化技术，遗传算法用于神经网络的训练就是很自然的事情。遗传算法用于神经网络的学习可分为三个不同的层次：连接权重的学习、网络拓扑的学习以及网络学习规则的学习。目前，遗传算法已广泛用于前向网络（feedforward networks）、径向基网络（radial basis function net-

works)、Kohonen 特征映射及 Recurrent 网络等各种人工神经网络的训练与设计中。Yao 在其两篇文章中对遗传算法在神经网络中的应用做了详细的综述，并把基于演化计算的神经网络系统（称作演化式神经网络，evolutionary artificial neural networks）作为一种一般的自适应学习模型加以研究。

被 Zadeh 称作软计算（soft computing）的两大组成部分——遗传算法与模糊系统的相互融合也是近年人们所关注的话题。模糊系统是对人类处理模糊性概念及其推理机制的模拟。最初，在模糊系统的设计中，推理方法的选取、隶属函数形状及参数的选取、相关权重的确定以及规则库的确定，均是由专家根据实际经验指定的，这也曾被认为是模糊系统较之经典解析模型的一大优点。但是人们也发现，对设计好的模糊系统中的一些参数基于数据进行调谐可大大提高模糊系统的性能，同时也有大量学者研究了基于经典优化的模糊系统训练学习，大部分可纳入模糊神经网络（fuzzy neural networks）框架。遗传算法同样能够用于模糊系统的设计中，并取得了十分显著的效果。遗传算法已成功应用于隶属函数形状与参数的优化，系统相关权重的优化以及推理规则的选取。此外，模糊集技术也被用于遗传算法的某些方面，形成称作模糊 GA 的新型遗传算法，以期达到改进经典遗传算法的目的。

大多数机器学习系统都有一个共同的特征，即具备对自身结构进行调整的能力，从而达到改进性能、发现并利用有意义的概念，或者改进其内部知识结构的一致性和通用性，Holland 最初引进遗传算法的目的也正是在自适应学习统计中设计一种有效的搜索机制。把遗传算法用于机器学习系统仍是目前遗传算法应用上十分活跃的领域。这一方面最著名的例子就是所谓的分类系统（classifer system），它是遗传算法与机器学习中经典的生产式系统（production system）相结合的产物，并成功应用于石油管道设计、视觉系统的控制以及布尔函数的推断。遗传算法也可用于有导师

概念学习(supervised concept learning)、特征选取与重构及强迫学习(reinforcement learning)等机器学习领域。

**2. 遗传算法设计与执行策略**

遗传算法操作的是一群编码化的可行解,称作种群。它通过种群的更新与迭代来搜索全局最优解。种群的迭代是通过选择、杂交和变异等具有生物意义的遗传算子来实现的。在 Holland 的最初模型中采用的是二进制定长编码和固定规模种群,遗传算法的主要形式为比例选择、单点杂交和位变异。为了提高遗传算法的性能,克服实际问题中遇到的困难,近年来在算法设计与执行策略方面有了很大进展。

(1) 编码方法　简单二进制编码的采用得到了 Holland 早期理论结果(Schema 定理、最小字母表原理)的支持,但它仍有许多不足之处。灰色编码可用于克服二进制编码映射的不连续问题(即欧氏空间中邻近点的二进制编码在 Hamming 距离下并不邻近)。动态参数编码(dynamic parameter encoding)的提出是为了克服搜索效率与表示精度间的矛盾,同时对克服过早收敛现象也有所帮助。此外,多值编码、实值编码、区间值编码、Delta 编码等多种编码方法也被证明各有优缺点。这些编码方法的提出是启发式的,缺乏一个理论基础来判断各种编码方法的好坏并指导它们的设计。

(2) 选择机制　选择是遗传算法中最主要的机制,也是影响遗传算法性能最主要的因素。选择压(selective pressure)描述了选择机制挑选种群中不同个体做母体的概率大小的差异。选择压过大,会造成几个较好可行解(不一定是近似全局最优解)迅速抢占了整个种群;选择压过小,则会使算法呈现出纯粹的随机徘徊行为。秩选择、适应值函数的尺度变换均是为了克服经典的比例选择所造成的选择压过大或过小而设计的。杰出者选择(elitist se-

lection)则通过无条件地保留每代种群中最优者而克服了比例选择下遗传算法不能依概率收敛到全局最优解的问题。然而正如De Jong 指出的，现有的遗传算法理论无法对这些现象做出完善的解释和提供指导。

(3) 杂交与变异机制 杂交与变异是遗传算法中最具争议性的两种机制。争议的焦点是：单点杂交、多点杂交与均匀杂交的优劣；杂交机制与变异机制的优劣。经典的观点强调杂交的作用，而认为变异只是一个背景机制，并且认为在杂交机制中强度最弱的单点杂交是最好的。这一观点与生物学中的实际观察是相符合的，但作为设计人工求解问题方法的思想，它正受到理论与实践两方面的挑战。

单点杂交优于多点、均匀杂交的观点是建立在 Holland 的 Schema 定理基础上的，但 Schema 定理中只考虑了杂交的破坏作用，而未考虑其建设性作用，因此推出“杂交强度应弱”的结论至少在逻辑上是不成立的(沿用同样的推理思路，甚至可以得出不要杂交机制是最好的)。同时，也有越来越多的实验结果显示强度较大的多点、均匀杂交优于单点杂交。

强调杂交作用的观点是基于遗传算法中著名的建筑块假设(building block hypothesis)，而这一假设并未得到理论证明。另外，在演化计算的另一分支——演化策略中并未用到任何形式的杂交机制，但并不妨碍它成为一种有效的优化技术。有鉴于此，尽管有大量为杂交机制辩护的结果，也有许多学者正从理论和实验两个方面探讨变异机制在遗传算法中所起的比杂交更显著、但又鲜为人知的作用。

此外，在杂交、变异概率的适应性调整，最优变异概率的确定，基于知识的启发式杂交机制的设计，种群规模的选取等方面，均有理论与实验上的进展。

(4) 执行策略的改进 除了关于世代型 GA(generational GA)

和稳定状态型 GA(steady state GA)的比较研究仍在进行以外,在执行策略方面的下列几方面工作值得注意:

a. 遗传算法与模拟退火算法的结合

模拟退火(simulated annealing)是一种基于热力学理论的优化方法,并有着完善的全局收敛理论。在遗传算法中以各种形式融入模拟退火思想,从而使得遗传算法在理论上具有全局收敛性,是一个有意义的研究方向。

b. 遗传算法与局部优化方法的结合

把遗传算法与局部搜索方法有机结合起来,是改进遗传算法性能的一个卓有成效的方法。这种混合型遗传算法不但模拟了生物种群的学习过程,而且事实上还模拟了种群的个体在其生命周期内具有学习行为这一生物现象。这在生物学中称作 lamarckian evolution 和 baldwin effect。

c. 并行遗传算法(parallel genetic algorithm)

并行遗传算法是在遗传算法的实施中加入自然生物种群中的空间结构因素,它的目的是可以在大规模并行机上运行,即使是序列式实施并行遗传算法,也可克服经典遗传算法性能上的不足,从而提高算法的效率。并行遗传算法大致分为两种:一种是岛模型(island model),称作粗粒并行遗传算法(coarsegrained PGA);另一种是细胞模型(cellular model),称为细粒并行遗传算法(fine-grained PGA)。岛模型的实施方法是在几个独立按经典 GA 方式运行的种群(称作子种群)中加入迁移(migration)因素;而细胞模型的实施则是在一个单一种群中,通过对经典整体性选择与杂交机制的局部化来达到并行化目的。岛模型考虑的是种群的空间结构,而细胞模型考虑的是一种群中个体的空间结构。

d. 共存演化遗传算法(cooperative coevolutionary GA, CCGA)

这一方法的思想是生物个体间不但有竞争,而且还具有合作

行为。CCGA 通过把复杂问题分解为数个子问题,通过代表这些子问题的个体间既合作又竞争的演化过程,达到求解问题的目的。

e. 混乱遗传算法(messy genetic algorithm)

混乱遗传算法是由 Goldberg 提出的,它的目的是解决经典 GA 中所谓的联结问题(linkage problem),即经典 GA 很容易破坏定义位置较远的模式的问题。在混乱遗传算法中,个体由其每个分量的位置与取值共同表示,位置可以不按次序排列,也并不是所有分量都在个体的表示中出现。事实上,混乱遗传算法直接操作的对象既有个体,也有 Schema,这自然使它丧失了只操作个体的经典 GA 所具有的隐含并行性。混乱遗传算法的理论基础、算法设计及实验比较,均是有待探讨的课题。

## 0-3 遗传算法的基础理论研究

遗传算法早期的基础理论主要是 Holland 的 Schema 定理及隐含并行性原理,后来又有了建筑块假设(building block hyothesis)。但是,Schema 定理却无法解释遗传算法实际操作中的许多现象,隐含并行性的论证存在严重的漏洞,而建筑块假设却从未得到过哪怕是启发式的证明。目前,关于遗传算法基础理论的研究在三个方面进行着,它们是:Schema 理论的拓广与深入;遗传算法的马氏链分析;遗传算法的收敛理论。

### 1. Schema 理论的拓广和深入

这一方面的工作主要包括 Schema 公式的进一步讨论与拓广。Radcliffe 在其一系列工作中,把 Schema 分析进行了一般化,提出完整的 forma 分析理论。在解决建筑块假设成立与否的问题上,目前的工作主要集中在所谓遗传算法的欺骗函数(decetive functions)的研究上。所谓欺骗函数就是那些对遗传算法进行误

导，使其错误地收敛到非全局最优解状态的函数。一旦研究清楚一个函数是遗传算法欺骗函数的条件，也就给出了建筑块假设成立的条件。研究欺骗函数问题的主要方法是 Walsh 变换。但对于确实有严重漏洞的隐含并行性原理，目前尚未有人提出改进办法，人们对这一遗传算法至关重要的优点只是加以主观信念上的默认与支持。

**2. 遗传算法的马氏链分析**

近年来，人们建立起了遗传算法不同形式的马氏链模型，对遗传算法的极限行为进行了各种角度的刻画。遗传算法的马氏链模型主要有三种，分别是种群马氏链模型、Vose 模型和 Cerf 扰动马氏链模型。

种群马氏链模型将遗传算法的种群迭代序列视为一个有限状态马氏链来加以研究。最早的工作属于 Goldberg。主要是运用种群马氏链转移概率矩阵的某些一般性质，分析遗传算法的极限行为，但转移概率的具体形式很难表达，这妨碍了对遗传算法有限时间行为的研究。

在 Vose 模型中，种群的状态由一个概率向量表示，概率向量的维数为所有可能个体的数目，第 $i$ 个分量表示第 $i$ 个个体在种群的个数比例(相对频率)。当种群规模趋于无穷大时，相对频率的极限就代表了每一个个体在种群中出现的概率。无限种群规模假设下，可以导出表示种群的概率向量的迭代方程。通过对这一迭代方程的研究，可以探讨种群概率向量的不动点及其稳定性，从而导致对遗传算法极限行为的刻画。虽然在无限种群假设下，Vose 模型可给出遗传算法极限行为的刻画，但它们解释实际有限种群遗传算法行为的能力相对差一些。另有一些工作则直接把种群相对频率序列当做一个马氏链加以研究。

法国学者 R. Cerf 在其一系列工作中，利用 Azencott, Catoni,

Trouve 等人关于模拟退火和广义模拟退火的一系列漂亮工作，将遗传算法看成一种特殊形式的广义模拟退火模型，利用动力系统的随机扰动理论，对遗传算法的极限行为及收敛速度进行了研究。尽管在 Cerf 模型中所研究的马氏链序列仍然是种群序列，但由于研究方法与种群马氏链模型的差异，我们将它称做 Cerf 扰动马氏链模型。

上述三种模型各有优缺点。种群马氏链模型最直观，因而对遗传算法行为的解释能力最强。但遗憾的是，由于对该马氏链转移概率没有一个深刻而细致的描述，目前所得结果仅仅用到了变异机制所导致的遍历性，因而只是形式上的；并且，所得到的算法收敛或不收敛结果的证明方法与纯随机抽样算法相应结果的证明方法在基本思想上无本质区别，并没有利用到遗传算法独有的特性。Vose 模型在理论上得出了一些形式复杂和漂亮的结果，但这些结果对遗传算法行为的解释性不强。Vose 模型的深入研究也许可以使遗传算法研究中用上已在群体遗传学上成功运用的随机分析方法。Cerf 的扰动马氏链模型得到了目前最完整的收敛性结果，而且有希望进一步深入。不足的是，它仍要假设种群规模趋于无穷大。

### 3. 遗传算法的收敛理论

对于遗传算法的马氏链分析本身，即是建立遗传算法的收敛性理论。早期的大多数文章，都是通过马氏链的极限理论分析遗传算法的收敛性。近几年来，我们的几位同事，如徐宗本教授、聂赞坎教授、高勇副教授、张讲社副教授等，在遗传算法收敛性方面进行了深入的研究，用鞅方法与迭代方法得到了比较系统的结果。

就遗传算法的目标来看，是寻求全局最优解，而完成的过程是一个随机搜索过程，总希望这个过程在期望值意义下越来越好，这样自然应当是一个下鞅序列，利用鞅收敛定理进一步证明遗传算

法的收敛性。同时也看到为了保证遗传算法的收敛性，有两个参数是非常重要的：一个参数是过程进入满意解后下一步脱离满意解集的可能性；另一个参数是过程未进入满意解时下一步仍不能进入满意解的可能性。这两个参数的匹配构成了遗传算法收敛的一般理论。用这种方法研究收敛性变为纯概率的方法，直观且简练地证明了多种遗传算法的收敛性。

遗传算法是一种仿生优化算法，是人类向其自身演化过程学习的结果。但是到目前为止，它始终未能对自身演化本身的研究发生作用。而且就其实质来讲，目前的遗传算法学习人类演化过程还只是形式的，尚未能刻画人自身的演化过程，更未能刻画神经元思维真实学习过程。因此遗传算法就其模型本身来讲需要更加深入的探讨。目前遗传算法的研究仅仅是一个开端，有必要从更高的高度与更加广阔的视野审视遗传算法的现状，并为未来探讨出一条新路。

# 第 1 章

# 遗传算法的几何理论

遗传算法是一种仿生算法，即模拟生命演化的算法。它是从一个初始种群出发，不断重复执行选择、杂交和变异的过程，使种群进化越来越接近某一目标。如果视种群为超空间的一组点，选择、杂交和变异的过程即是在超空间中进行点集之间的某种变换，通过信息交换使种群不断进化。

## 1-1 遗传算法的基本概念

遗传算法被认为是对人类自然演化过程的模拟。人类的自然演化过程是进化过程，这种进化过程发生在染色体上；自然选择使适应值好的染色体比那些适应值差的染色体有更多的繁殖机会；变异可以使子代染色体不同于父代染色体；通过两个父代染色体的结合与重组可以产生全新的染色体。染色体的选择、变异与重组

进程是无记忆的。将这些概念反映在数学上就形成了遗传算法的基本概念。

**定义 1.1.1**(个体和个体空间)　所谓 $l$- 个体 $X$,即是长度为 $l$ 的 0 和 1 字符串,简称个体;$l$ 称作个体的链长,$l$- 个体的全体记作 $S = \{0,1\}^l$,称为个体空间。

**例 1.1.1**　对于一个饭店的经营决策,考虑三个因素:价格、饮料与服务速度。三个因素称为三个变量。对于价格,用"0"表示价格高,用 1 表示价格低;对于饮料,用"0"表示啤酒,用"1"表示可乐;对于服务速度,用"0"表示"慢",用 1 表示快。那么

$$X = (0\ 1\ 1)$$

表示个体,个体的链长 $l = 3$, $S = \{0,1\}^3$ 有 8 个个体。$X = (0\ 1\ 1)$ 表示采用高价格、可乐饮料和服务速度快的经营决策。

按照遗传学的术语,个体也称作染色体(chromosome),个体的分量称作基因(gene),分量的取值称作等位基因(allele)。

**例 1.1.2**(函数优化问题)　设 $F(x)$ 是 $n$ 维实数空间 $\mathbf{R}^n$ 到正实数 $\mathbf{R}^+$ 的函数。对于 $x \in \mathbf{R}^n$ 可与一个长度为 $l$ 的二进制对应,于是求解 $\max F(x)$ 的问题可以转化为求解

$$\max\{f(X);\ X \in \{0,1\}^l\}$$

其中

$$f(X) = S(F(e^{-1}(X)))$$

$e$ 是 $\mathbf{R}^n$ 到 $\{0,1\}^l$ 的编码映射,$S$ 是 $\mathbf{R}^+$ 到 $\mathbf{R}^+$ 的增函数。通常采用的二进制编码映射为固定长度 $l$ 的二进制编码。二进制编码称为个体,长度 $l$ 的二进制编码全体称为个体空间。二进制编码可以是实数的二进制表示,也可以是按照某种规则进行的二进制编码,但是要求二进制编码与实数之间一一对应。例如在 $[0,31]$ 的整数上求 $F(x) = x^2$ 的最大值,可直接利用实数的二进制表示,$X = (1\ 0\ 1\ 1\ 1)$ 表示 $16 + 4 + 2 + 1 = 23$。问题转化为求解

$$\max\{f(X);\ X \in \{0,1\}^l\}$$

$$f(X) = F(e^{-1}(X))$$

其中 $e$ 为二进制编码方式。

一般地，一个个体的分量可以在 $\{0,1,\cdots,k\}$ 中取值，这时个体空间为

$$S = \{0,1,\cdots,k\}^l$$

**定义 1.1.2**(种群和种群空间)　所谓 $N$- 种群是 $N$ 个个体组成的集合(个体允许重复)，简称种群。$N$ 称作种群规模，称

$$S^N = \{\vec{X} = (X_1,X_2,\cdots,X_N),\ X_i \in S(i \leqslant N)\}$$

为 $N$- 种群空间。

**例 1.1.3**(续例 1.1.1)　对于一个饭店的经营决策，一个个体表示一种经营决策，则

$$X_1 = (0\ 1\ 1)$$

$$X_2 = (0\ 0\ 1)$$

$$X_3 = (1\ 1\ 0)$$

$$X_4 = (0\ 1\ 0)$$

表示 4 种经营决策。于是

$$\vec{X}_1 = (X_1,X_2,X_3,\ X_4)$$

为种群规模 $N = 4$ 的种群。

对于一个种群，它的个体可以是相同的，例如

$$\vec{X}_2 = (X_1,X_2,X_2,\ X_1)$$

也是一个种群规模 $N = 4$ 的种群，但实际上它只有两个不重复的个体。

对于一个种群，由于是由个体构成的，因此它是个体空间的一部分，即是个体空间的一个子集。我们可以记为 $\vec{X}_1 \subset S$，$\vec{X}_2 \subset S$，当用 $\vec{X}_1 \subset S$ 时，视 $\vec{X}_1$ 为一个子集，当用 $\vec{X}_2 \subset S$ 时，表示 $\vec{X}_2$ 中不重复的个体是 $S$ 中的一个子集。这时 $\vec{X}_1$ 和 $\vec{X}_2$ 都不视为向量。如果视 $\vec{X}$ 为向量，则 $\vec{X}$ 可表示为矩阵形式。如：

$$\vec{X}_1 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad \vec{X}_2 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

其中每一行代表一个个体，第 $i$ 行表示第 $i$ 个个体。

对于链长为 $l$ 种群规模为 $N$ 的种群可以表示为 $N \times l$ 阶矩阵：

$$\vec{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1l} \\ x_{21} & x_{22} & \cdots & x_{2l} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nl} \end{pmatrix}$$

其中

$$X_i = (x_{i1}\ x_{i2} \cdots x_{il})$$

为一个个体。$x_{ij}$ 表示第 $i$ 个个体第 $j$ 个基因取值。

**定义 1.1.3**(齐次种群)　称 $N$- 种群是齐次种群，若对于任意 $i,j \leqslant N$ 有 $X_i = X_j$，用 $\beta^N$ 表示齐次种群的全体。

显然 $\beta^N \subset S^N$。对于齐次种群，每一个齐次种群仅有一个重复的个体。因此，$\beta^N$ 中的数目应与个体空间 $S$ 中的数目相同，即

$$|\beta^N| = |S|$$

其中 $|\cdot|$ 表示集合中元素的个数。

对于链长 $l$ 的个体空间，

$$|S| = 2^l \qquad |S^N| = 2^{Nl}$$

$$|\beta^N| = |S| = 2^l$$

例如在例 1.1.1 中

$$|S| = |\beta^N| = 2^3 = 8$$

在例 1.1.3 中

$$|S^N| = 2^{4\times 3} = 4\ 096$$

**定义 1.1.4**　对于 $\vec{X} \in S^N$，$X \in S$，记

$$\varphi(\vec{X},X)=\sum_{i=1}^{N}\chi_{\{X\}}(X_i)$$

其中 $\chi_A$ 表示$A$的示性函数，$\varphi(\vec{X},X)$ 表示种群 $\vec{X}$ 包含个体$X$的数目。称两个种群 $\vec{X}$ 和 $\vec{Y}$ 是等价的，若对于任意 $X\in S$ 有

$$\varphi(\vec{X},X)=\varphi(\vec{Y},X)$$

记作 $\vec{X}\sim\vec{Y}$。

$\vec{X}\sim\vec{Y}$ 表示两层意思：一是 $\vec{X}$ 与 $\vec{Y}$ 中含有相同的个体；二是 $\vec{X}$ 与 $\vec{Y}$ 中含有的相同的个体重复的次数一样多。因此，$\vec{X}\sim\vec{Y}$ 表示 $\vec{Y}$ 是 $\vec{X}$ 中的个体次序的某种变换。

**定理 1.1.1**　“$\sim$”是 $S^N$ 上的等价关系。

**证明**　首先易证 $\vec{X}\sim\vec{X}$。若 $\vec{X}\sim\vec{Y}$，则 $\vec{Y}\sim\vec{X}$。如果 $\vec{X}\sim\vec{Y}$，$\vec{Y}\sim\vec{Z}$，则对于任意 $X\in S$ 有

$$\varphi(\vec{X},X)=\varphi(\vec{Y},X)=\varphi(\vec{Z},X)$$

于是 $\vec{X}\sim\vec{Z}$。

**定理 1.1.2**　等价关系“$\sim$”在 $S^N$ 上诱导的等价类划分 $\mathcal{Q}=S^N/\sim$ 的个数为

$$\mathcal{Q}=\binom{N+2^l-1}{2^l-1}$$

**证明**　根据定义 1.1.4 等价类划分 $\mathcal{Q}$ 与集合

$$\Phi=\left\{(\varphi_1,\varphi_2,\cdots,\varphi_{2^l});\varphi_i\ \text{为非负整数},\sum_{i=1}^{2^l}\varphi_i=N\right\}$$

一一对应。根据组合学原理则证 $|\Phi|=\mathcal{Q}$。

考虑例 1.1.1 中 $l=3$，$N=4$，于是种群空间的个数为 $2^{3\times4}=2^{12}=4\,096$，个体空间为 $2^3=8$，而等价类空间 $\mathcal{Q}$ 为

$$\mathcal{Q}=\binom{4+8-1}{8-1}=\binom{11}{7}$$

$$=\frac{11!}{7!4!}=\frac{11\times10\times9\times8}{4!}=330$$

这样等价类种群个数为 330，与种群空间个数 4 096 比较起来还不到 1/10。

我们在使用种群时只关心种群中的个体的作用，这些个体出现次数在某种程度上反映了它的性能，而这些个体如何排列是不起作用的。因此，使用等价类个体可以使遗传过程简单化。

遗传算法的基本概念构成了遗传算法的基本模型。对于一个实际问题，首先是将问题转化为遗传算法的基本概念。一旦用遗传算法的基本概念来表达实际问题，问题的求解过程就不再依赖于实际问题本身。

**例 1.1.4**(Goldberg, Samtani)10 杆桁架负载问题。

图 1.1.1 给出了 10 杆桁架的结构示意图。其中 10 个桁架的截面面积分别表示为 $A_1, A_2, \cdots, A_{10}$。这个桁架由左边的墙支撑，并且它必须能承受图中所示的两个负载。每个杆上的应力必须在一个允许值范围内，这个范围由那个杆的应力约束来表示。优化的目标是找到这个负载桁架的每个杆的截面面积以使建造它的材料总重量最小。

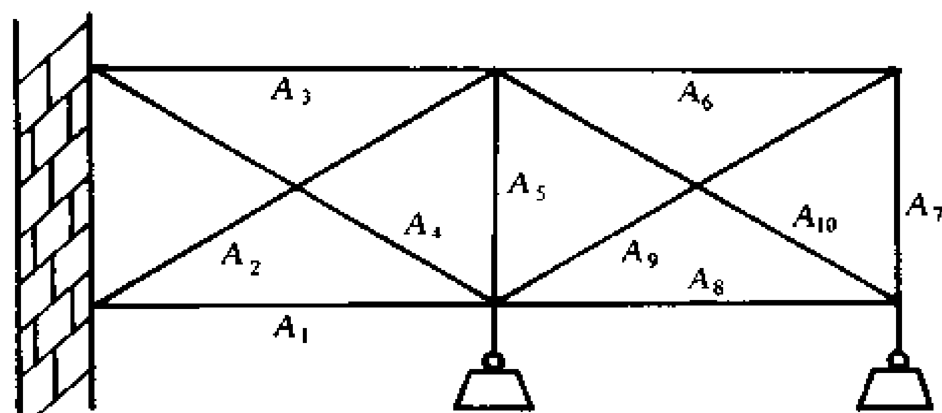



图 1.1.1　10 杆桁架

下面我们用遗传算法的基本概念来表示这个问题。首先假定截面面积在 0.1 ～ 10 面积单位之间有 16 个可能的值。其中等于

0.1面积单位的截面面积用4位数串(0000)表示，等于10面积单位的截面面积用4位数串(1111)表示，其它的具有中间值的截面面积用余下的14个4位数串编码，即(0001),(0010),(0011),…,(1110)等。对10个桁架$A_1,\cdots,A_{10}$给出10个截面面积，即构成一个个体。如

$$X_1 = (0010111000010011101100111111001100111010)$$

即是一个个体，个体的链长为40。于是个体空间为

$$S = \{0,1\}^{40}$$

个体空间中有$2^{40} \sim 10^{12}$个个体。

如果取$N = 200$，即种群规模为200，则种群空间为：

$$S^{200} = \{0,1\}^{40\times 200}$$

一个种群可以表示为

$$\vec{X} = (X_1, X_2, \cdots, X_{200})$$

$\vec{X}$代表着$200\times 40$阶的矩阵，矩阵的每一行代表一个个体，第$i$行代表第$i$个个体。如果第$i$个个体表示为

$$X_i = (x_{i1}, x_{i2}, \cdots, x_{i,40})$$

则$x_{ij}$是一个基因，它可以取1或0。前4个基因位量表示桁架$A_1$的截面积，$(x_{i5}, x_{i6}, x_{i7}, x_{i8})$表示第二个桁架的截面积，依次类推。每4个基因构成一个整体，它的变化直接表示某一桁杆截面面积的变化。每个桁杆有一个应力约束，如果违犯了应力约束，则这个解是不可行解。

Goldberg和Samtani于1986年在第九届电子计算机国际会议上所发表的论文，建立了以上问题的遗传算法模拟。通过遗传算法40代的计算得到了10杆桁架问题的可行设计，且其材料总费用不超过理论最优解的1%。

## 1-2 遗传机制与遗传算法

遗传机制主要包括选择、杂交和变异。遗传算法是对种群通过遗传机制模拟自然进化过程而不断产生新一代种群，并使新一代种群有更好的性质。

**定义 1.2.1**(适应值函数) 对于一个个体产生的效益称为适应值。适应值函数是个体空间 $S$ 到正实数空间的映射，即适应值函数 $f$ 为

$$f: S \to \mathbf{R}^1$$

在例 1.1.1 中，一个个体的适应值即是一个经营决策产生的利润。在例 1.1.2 中，一个个体的适应值即是个体所代表的实数的函数值。例 1.1.1 与例 1.1.2 恰好反映两大类问题，即决策问题与优化问题。

**定义 1.2.2**(选择算子) 选择算子即是在一个种群中选择一个个体，它是随机映射：

$$T_s: S^N \to S$$

特别地，按照概率规则

$$P\{T_s^\alpha(\vec{X}) = X_i\} = \frac{f^\alpha(X_i)}{\sum_{k=1}^{N} f^\alpha(X_k)} \tag{1.2.1}$$

选择个体的方式称为适应值选择算子。其中 $f(X_i)$ 表示种群 $\vec{X}$ 中的个体 $X_i$ 的适应值，$0 < \alpha < \infty$。对于 $\alpha = 1$ 称为比例选择算子，简记为 $T_s$。

容易看到适应值选择算子(依赖于种群 $\vec{X}$，它是种群 $\vec{X}$ 上的一个概率分布，即对于种群

$$\vec{X} = (X_1, X_2, \cdots, X_N)$$

有概率分布

$$P\{X_i\} = f^{\alpha}(X_i) / \sum_{k=1}^{N} f^{\alpha}(X_k)$$

即构成概率分布向量

$$\begin{bmatrix} X_1 & X_2 & \cdots & X_N \\ P\{X_1\} & P\{X_2\} & \cdots & P\{X_N\} \end{bmatrix}$$

其中 $P\{X_i\}$ 满足以下条件：

(1) $P\{X_i\} > 0 \ (i \leqslant N)$

(2) $\sum_{i=1}^{N} P\{X_i\} = 1$

称 $P\{X_i\}(i \leqslant N)$ 为 $\vec{X}$ 上的一维概率分布。

**例 1.2.1**(续例 1.1.3)　给出饭店经营决策的初始种群与适应值

$$X_1 = (0\ 1\ 1) \qquad f(X_1) = 3$$
$$X_2 = (0\ 0\ 1) \qquad f(X_2) = 1$$
$$X_3 = (1\ 1\ 0) \qquad f(X_3) = 6$$
$$X_4 = (0\ 1\ 0) \qquad f(X_4) = 2$$

采用比例选择算子 $T_s$，则得到

$$\frac{f(X_1)}{\sum_{k=1}^{4} f(X_k)} = \frac{3}{12} = \frac{1}{4}, \quad \frac{f(X_2)}{\sum_{k=1}^{4} f(X_k)} = \frac{1}{12}$$

$$\frac{f(X_3)}{\sum_{k=1}^{4} f(X_k)} = \frac{6}{12} = \frac{1}{2}, \quad \frac{f(X_4)}{\sum_{k=1}^{4} f(X_k)} = \frac{1}{6}$$

于是得到初始种群的分布

$$\begin{bmatrix} X_1 & X_2 & X_3 & X_4 \\ 0.25 & 0.08 & 0.5 & 0.17 \end{bmatrix}$$

按照概率规则选择个体 $X_1$ 和 $X_4$ 易保留一次，$X_2$ 可能被淘汰，而 $X_3$ 可能被选择两次，这样选择以后的结果可能为 $\vec{X} = (X_1, X_3,$

$X_3, X_4$)。

按照式(1.2.1)计算即得到新的概率分布为

$$\begin{bmatrix} X_1 & X_3 & X_3 & X_4 \\ 0.18 & 0.35 & 0.35 & 0.12 \end{bmatrix}$$

从选择算子的执行结果可以看到，适应值大的个体易被选择，适应值小的个体易被淘汰，这样经过不断地选择使适应值大的个体不断重复出现，使一个种群可能变为齐次种群。但是选择算子只是在一个固定的种群中进行选择，选择的结果不可能跑到种群之外。对于种群之外更好的个体是不可能被选择得到的，选择的结果依赖于初始种群。

**定义 1.2.3**(母体和母体空间)　所谓母体即是一对个体 $(X_1, X_2)$，其中 $X_i \in S(i = 1,2)$。所有母体的全体称为母体空间，即

$$S^2 = \{(X_1, X_2);\ X_1, X_2 \in S\}$$

**定义 1.2.4**(母体选择算子)　母体选择算子即是在一个种群中选择一个母体，它是映射：

$$T_s:\ S^N \to S^2$$

特别地，按照适应值选择算子独立的选择两个个体而构成母体的方式为适应值母体选择算子，即

$$\begin{aligned} P\{X_i, X_j\} &= P\{T_s^\alpha(\vec{X}) = (X_i, X_j)\} \\ &= \frac{f^\alpha(X_i)}{\sum_{k=1}^{N} f^\alpha(X_k)} \cdot \frac{f^\alpha(X_j)}{\sum_{l=1}^{N} f^\alpha(X_l)} \end{aligned} \tag{1.2.2}$$

容易看到母体选择算子构成种群乘积空间 $\vec{X} \otimes \vec{X}$ 上的二维概率分布

$$P = \begin{bmatrix} (X_i, X_j) & \\ P\{X_i, X_j\} & i, j \leqslant N \end{bmatrix}$$

二维概率分布 $P$ 是 $N \times N$ 阶矩阵，且满足性质：

(1) $P\{X_i, X_j\} > 0, (i, j \leqslant N)$

(2) $\sum_{i=1}^{N}\sum_{j=1}^{N} P\{X_i, X_j\} = 1$

它与种群 $\vec{X}$ 上的概率分布有关系:

(1) $\sum_{i=1}^{N} P\{X_i, X_j\} = P\{X_j\}$

(2) $\sum_{j=1}^{N} P\{X_i, X_j\} = P\{X_i\}$

亦即一维概率分布是二维概率分布的边际分布。

**例 1.2.2**(续例 1.2.1)　考虑例 1.2.1 中的 4 个个体和它的适应值,于是得到

$$P\{T_s(\vec{X}) = (X_1, X_1)\} = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$$

$$P\{T_s(\vec{X}) = (X_1, X_2)\} = \frac{1}{4} \times \frac{1}{12} = \frac{1}{48}$$

$$\vdots$$

从而得到$(X_i, X_j)(i, j \leqslant 4)$ 的二维概率分布为

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{16} & \frac{1}{48} & \frac{1}{8} & \frac{1}{24} \\ \frac{1}{48} & \frac{1}{144} & \frac{1}{24} & \frac{1}{36} \\ \frac{1}{8} & \frac{1}{24} & \frac{1}{4} & \frac{1}{12} \\ \frac{1}{24} & \frac{1}{36} & \frac{1}{12} & \frac{1}{36} \end{bmatrix}$$

**定理 1.2.1**　对于选择算子 $T_s^r$ 及种群

$$\vec{X} = (X_1, X_2, \cdots, X_N)$$

有性质:

(1) $P\{T_s^\alpha(\vec{X})\subset\vec{X}\}=1$,即 $T_s^\alpha(\vec{X})\subset\vec{X}$

(2) $P\{T_s^\alpha(\vec{X})=X\}=\dfrac{\varphi(\vec{X},X)f^\alpha(X)}{\sum\limits_{k=1}^{N}f^\alpha(X_k)}$

(3) $P\{T_s^\alpha(\vec{X})=(X,Y)\}=\dfrac{\varphi(\vec{X},X)f^\alpha(X)}{\sum\limits_{k=1}^{N}f^\alpha(X_k)}\,\dfrac{\varphi(\vec{X},Y)f^\alpha(Y)}{\sum\limits_{k=1}^{N}f^\alpha(X_k)}$

(1.2.3)

**证明** (1) 由于

$$P\{T_s^\alpha(\vec{X})\subset\vec{X}\}=\sum_{i=1}^{N}P\{T_s^\alpha(\vec{X})=X_i\}=1$$

则证(1)。

(2) 对于 $X\in S$,由全概率公式

$$\begin{aligned}
&P\{T_s^\alpha(\vec{X})=X\}\\
&=\sum_{i=1}^{N}P\{T_s^\alpha(\vec{X})=X/X=X_i\}P\{X=X_i\}\\
&=\sum_{i=1}^{N}P\{T_s^\alpha(\vec{X})=X_i\}\chi_{\{X_i\}}(X)\\
&=\frac{\varphi(\vec{X},X)\cdot f^\alpha(X)}{\sum\limits_{i=1}^{N}f^\alpha(X_i)}
\end{aligned}$$

(3) 类似(2)可证。

**推论 1.2.1** 若 $\vec{X}\sim\vec{Y}$,则 $T_s^\alpha(\vec{X})$ 与 $T_s^\alpha(\vec{Y})$ 同分布。

**证明** 由于 $\vec{X}\sim\vec{Y}$,则 $\varphi(\vec{X},X)=\varphi(\vec{Y},X)$ 对于任意 $X\in S$ 成立,据定理 1.2.1 则证。

**推论 1.2.2** 若 $\vec{X}=(X,X,\cdots,X)\in\beta^N$ 为齐次种群,则

$$P\{T_s^\alpha(\vec{X})=X\}=1$$

$$P\{T_s^\alpha(\vec{X})=(X,X)\}=1$$

**证明** 由于 $\varphi(\vec{X},X)=N$,且

$$\sum_{h=1}^{N} f^{\alpha}(X_h) = Nf^{\alpha}(X)$$

据定理 1.2.1 则证。

**定理 1.2.2** (1) 设 $\vec{X}$ 为一种群，$I \subset \{1,2,\cdots,N\}$ 使得 $f(X_i) = f(X_j)(i,j \in I)$，且 $f(X_i) > f(X_j)(i \in I, j \notin I)$，则

$$\lim_{\alpha\to\infty} P\{T_s^{\alpha}(\vec{X}) = (X_i, X_j)\} = \frac{1}{K^2} \qquad (i,j \in I)$$

其中 $K = |I|$ 为 $I$ 中元素个数。

(2) 任给种群 $\vec{X}$，必有

$$\lim_{\alpha\to 0} P\{T_s^{\alpha}(\vec{X}) = (X_i, X_j)\} = \frac{1}{N^2} \qquad (i,j \leqslant N)$$

**证明**　(1) 依定理假设及式(1.2.2) 有

$$\begin{aligned}
&\lim_{\alpha\to\infty} P\{T_s^{\alpha}(\vec{X}) = (X_i, X_j)\} \\
&= \lim_{\alpha\to\infty}\left[\frac{\sum_{k=1}^{N} f^{\alpha}(X_k)}{f^{\alpha}(X_i)}\right]^{-1}\left[\frac{\sum_{k=1}^{N} f^{\alpha}(X_k)}{f^{\alpha}(X_j)}\right]^{-1} \\
&= \lim_{\alpha\to\infty}\left[K + \sum_{k\notin I}\frac{f^{\alpha}(X_k)}{f^{\alpha}(X_i)}\right]^{-1}\left[K + \sum_{k\notin I}\frac{f^{\alpha}(X_k)}{f^{\alpha}(X_j)}\right]^{-1} \\
&= \frac{1}{K^2}
\end{aligned}$$

(2) 类似(1) 可证。

由定理 1.2.2 可知，选择算子的指数 $\alpha$ 对于选择有重要作用。$\alpha$ 越大，适应值大的两个个体越容易被选择；$\alpha$ 越小，选择的随机性越大，适应值大小对选择不起作用。$\alpha$ 很大时搜索空间很小，基本上是在一个种群中适应值大的个体中搜索。而当 $\alpha$ 很小时搜索空间很大，它在种群的全体中搜索。选择算子具有“优胜”的性质。

**定义 1.2.5**(杂交算子)　杂交算子是母体空间到个体空间的映射，记作

$T_c: S^2 \to S$

(1) 单点杂交算子：等概率的随机确定一个基因位置作为杂交点，再把一对母体两个个体从杂交点分成前后两个部分，交换两个个体的后半部分得到两个新个体，取第一个个体为杂交结果。

(2) 单点随机杂交算子：等概率的随机确定一个基因位置作为杂交点，再把一对母体两个个体从杂交点分为前后两个部分，以概率 $p_c$ 交换两个个体的后半部分，得到两个新个体，取第一个个体为杂交结果。称 $p_c$ 为杂交概率。

如果确定两个基因位置将一对母体两个个体分成三部分，交换中间部分，称为双点杂交算子。同样有双点杂交与双点随机杂交之分。

(3) 均匀随机杂交算子：独立地以概率 $p_c$ 把母体的第一个个体的相应的分量交换为第二母体的相应分量，从而得到杂交结果。

**例 1.2.3**(续例 1.2.1)　取两个个体

$$X_3 = (1\ 1\ 0) \qquad X_1 = (0\ 1\ 1)$$

假定第 2 位为杂交点，则将 $X_1$ 和 $X_3$ 的最后一位相互交换得到

$$X_3' = (1\ 1\ 1) \qquad X_1' = (0\ 1\ 0)$$

于是 $X_3$ 和 $X_1$ 杂交后得到新的个体 $X_3' = (1\ 1\ 1)$。这就是单点杂交算子杂交的结果。如果 $p_c = 0.5$，则各有 0.5 的概率使杂交结果为(1 1 1)，或保留(1 1 0)。

如果采用均匀随机杂交算子则可能有以下情形：

$$(1\ 1\ 0)\quad (0\ 1\ 0)\quad (1\ 1\ 1)\quad (0\ 1\ 1)$$

由例 1.2.3 可见，杂交算子的结果可能在种群之中，也可能在种群之外，这一点是与选择算子不同的。杂交算子扩大了搜索空间。

**定理 1.2.3**　(1) 设 $T_c$ 为单点杂交算子，对于任意$(X_1, X_2) \in S^2$，$Y \in S$，则

$$P\{T_c(X_1,X_2)=Y\}=\frac{k}{l} \tag{1.2.4}$$

其中 $k=k(X_1,X_2,Y)$ 为用单点杂交$(X_1,X_2)$可以生成 $Y$ 的基因位置的个数。

(2) 设 $T_c$ 为单点随机杂交算子，$p_c$ 为杂交概率，对于任意$(X_1,X_2)\in S^2$，$Y\in S$，则

$$P\{T_c(X_1,X_2)=Y\}=\begin{cases}\dfrac{kp_c}{l}, & Y\neq X_1\\(1-p_c)+\dfrac{kp_c}{l}, & Y=X_1\end{cases} \tag{1.2.5}$$

(3) 设 $T_c$ 为均匀随机杂交算子，对于任意$(X_1,X_2)\in S^2$，$Y\in S$，$X_i=(x_{i1},\cdots,x_{il})(i=1,2)$，$Y=(y_1,\cdots,y_l)$，有

$$P\{T_c(X_1,X_2)=Y\}=\prod_{j=1}^{l}[(1-p_c)\cdot(y_j\oplus x_{1j})+p_c\cdot(y_j\oplus x_{2j})] \tag{1.2.6}$$

其中 $\oplus$ 为异或算子，即对于任意 $x,y\in\{0,1\}$ 有

$$x\oplus y=\begin{cases}1, & x=y\\0, & x\neq y\end{cases}$$

**证明**　(1) 由于杂交点有 $l$ 种选择 $0,1,2,\cdots,l-1$，每一种杂交点的选择通过单点杂交后可能是 $Y$，也可能不是 $Y$。$k$ 表示了选择的杂交点通过杂交后生成 $Y$ 的个数，而杂交点的选择是均匀分布的，则证(1)。

(2) 单点随机杂交算子分为两种情形考虑。在 $Y\neq X_1$ 时，取决于杂交$(X_1,X_2)$可以生成 $Y$ 的基因位置的个数 $k$，与(1)不同的是即使选择的杂交点杂交后可以生成 $Y$，而它未必杂交，杂交的概率为 $p_c$。于是 $Y\neq X_1$ 时

$$P\{T_c(X_1,X_2)=Y\}=\frac{kp_c}{l}$$

当 $Y=X_1$ 时，$X_1$ 与 $X_2$ 在任何杂交点都不杂交即为 $Y=X_1$，这时概率为 $(1-p_c)$，$X_1$ 与 $X_2$ 在任何杂交点以 $p_c$ 概率杂交，这时概率为 $\frac{kp_c}{l}$，根据概率加法公式有

$$P\{T_c(X_1,X_2)=Y\}=(1-p_c)+p_c\cdot\frac{k}{l}$$

则证(2)。

(3) 设 $T_c$ 为均匀随机杂交算子，记：

$$X_1=(x_{11},x_{12},\cdots,x_{1l})$$
$$X_2=(x_{21},x_{22},\cdots,x_{2l})$$

考虑 $Y$ 的第 $j$ 个分量，若 $x_{1j}$ 与 $y_j$ 相同，这时不杂交；而当 $x_{2j}$ 与 $y_j$ 相同时，这时杂交。于是利用全概率公式

$$\begin{aligned}P\{T_c(X_1,X_2)_j=y_j\}&=P\{T_c(X_1,X_2)_j=y_j/y_j=x_{1j}\}P\{y_j=x_{1j}\}\\&\quad+P\{T_c(X_1,X_2)_j=y_j/y_j=x_{2j}\}P\{y_j=x_{2j}\}\\&=(1-p_c)\cdot(y_j\oplus' x_{1j})+p_c\cdot(y_j\oplus' x_{2j})\end{aligned}$$

由于均匀随机杂交算子各分量杂交是独立实现的，则证(3)。

**推论 1.2.3** 设 $T_c$ 为杂交算子(单点杂交，单点随机杂交、均匀随机杂交)，对于任何母体 $(X,X)$，必有

$$P\{T_c(X,X)=X\}=1$$

即

$$T_c(X,X)=X$$

**证明** 当 $T_c$ 为单点杂交时，$k(X,X,X)=l$，由定理 1.2.3 的(1)与(2)则证。当 $T_c$ 为均匀随机杂交时，由于对于任何 $j\leqslant l$ 有

$$x_j \oplus \bar{x}_j = 1$$

依定理 1.2.3(3) 则证。

**定理 1.2.4**　设 $T_c$ 为单点随机杂交算子或均匀随机杂交算子，对于任给种群 $\vec{X} = (X_1, X_2, \cdots, X_N)$，以及 $i \leqslant N$，必有

$$P\{T_c(T_s^\alpha(\vec{X})) = X_i\} > 0 \tag{1.2.7}$$

**证明**　首先

$$\begin{aligned} &P\{T_c(T_s^\alpha(\vec{X})) = X_i\} \\ &\quad \geqslant P\{T_c(X_i, X_i) = X_i\} \cdot P\{T_s^\alpha(\vec{X}) = (X_i, X_i)\} \end{aligned}$$

又因为

$$P\{T_s^\alpha(\vec{X}) = (X_i, X_i)\} = \left(\frac{f^\alpha(X_i)}{\sum_{k=1}^{N} f^\alpha(X_k)}\right)^2 > 0$$

而 $P\{T_c(X_i, X_i) = X_i\} = 1$，则证。

**推论 1.2.4**　设 $T_c$ 为杂交算子，$T_s^\alpha$ 为选择算子，$\vec{X} = (X, \cdots, X)$ 为齐次种群，则

$$P\{T_c(T_s^\alpha(\vec{X})) = X\} = 1$$

**证明**　由于

$$\begin{aligned} &P\{T_c(T_s^\alpha(\vec{X})) = X\} \\ &\quad = P\{T_s^\alpha(\vec{X}) = (X, X)\} P\{T_c(X, X) = X\} \end{aligned}$$

利用推论 1.2.2 和推论 1.2.3 则证。

**定义 1.2.6**(变异算子)　变异算子是个体空间到个体空间的随机映射 $T_m: S \to S$，其作用方式为独立地以概率 $p_m$ 改变个体分量取值。$p_m$ 称作变异概率。

**定理 1.2.5**　任给两个个体 $X, Y \in S$，有

$$P\{T_m(X) = Y\} = p_m^{d(X,Y)}(1 - p_m)^{l - d(X,Y)} \tag{1.2.8}$$

其中 $d(X, Y)$ 表示 $X$ 与 $Y$ 之间的 Hamming 距离，即

$$d(X, Y) = \sum_{i=1}^{l} | x_i - y_i |$$

其中

$$X=(x_1,x_2,\cdots,x_l)$$

$$Y=(y_1,y_2,\cdots,y_l)$$

**证明** 将 $Y$ 的分量分为两种，一种是 $y_i\neq x_i$，一种是 $y_i=x_i$。显然 $y_i\neq x_i$ 的个数为 $d(X,Y)$，这时每个 $x_i$ 变异 $y_i$ 的概率为 $p_m$。$y_i=x_i$ 的个数为 $l-d(X,Y)$，这时 $x_i$ 不变异，而不变异的概率为 $(1-p_m)$。由于各个基因的变异是相互独立的，则证(1.2.8)成立。

基于生物学上的考虑，一般认为杂交是自然演化的主要机制，变异为自然演化的背景，它们分别承担遗传与变异两种功能。因此在具体应用过程中，杂交概率一般取值较大，在 0.65 与 0.9 之间。而变异概率取值较小，一般在 0.001 与 0.01 之间。

**定理 1.2.6** 设 $T_s^a$ 为选择算子，$T_c$ 为杂交算子，$T_m$ 为变异算子，则对于任意 $\vec{X}\in S^N$，$Y\in S$，使

$$P\{T_m\circ T_c\circ T_s^a(\vec{X})=Y\}>0$$

的充要条件为 $p_m$ 满足 $0<p_m<1$。其中

$$T_m\circ T_c\circ T_s^a(\vec{X})=T_m(T_c(T_s^a(\vec{X})))$$

**证明** 充分性： 当 $0<p_m<1$ 时，由定理 1.2.5 知对于任意 $X,Y\in S$ 有 $P\{T_m(X)=Y\}>0$。因此，对于任意 $\vec{X}\in S^N$，$Y\in S$，根据定理 1.2.4 有

$$P\{T_m\circ T_c\circ T_s^a(\vec{X})=Y\}$$

$$=\sum_{X\in S}P\{T_m(X)=Y\}\cdot P\{T_c\circ T_s^a(\vec{X})=X\}>0$$

必要性： 假设 $p_m=0$。设 $\vec{X}=(X_0,\cdots,X_0)$ 为齐次种群，$Y\in S$，$Y\neq X_0$。由于

$$P\{T_c\circ T_s^a(\vec{X})=X_0\}=1$$

$$P\{T_m(X_0)=Y\}=0$$

于是由全概率公式有

$$\begin{aligned}
&P\{T_m \circ T_c \circ T_s^a(\vec{X}) = Y\} \\
&= \sum_{X\in S} P\{T_m(X) = Y\} \cdot P\{T_c \circ T_s^a(\vec{X}) = X\} \\
&= P\{T_m(X_0) = Y\} \cdot P\{T_c \circ T_s^a(\vec{X}) = X_0\} = 0
\end{aligned}$$

与 $P\{T_m \circ T_c \circ T_s^a(\vec{X}) = Y\} > 0$ 假定矛盾，故 $p_m > 0$。同理可证 $p_m < 1$。

**定义 1.2.7**(删除算子)　删除算子是从一个种群删除掉一个个体，它是 $S^N$ 到 $S^{N-1}$ 的随机映射，即

$$T_d: S^N \to S^{N-1}$$

是指数为 $0 \leqslant \alpha \leqslant \infty$ 的随机删除算子，由下面的概率分布确定：

$$P\{T_d^a(\vec{X}) = \vec{X} \setminus \{X_i\}\} = \frac{f^{-a}(X_i)}{\sum_{j=1}^{N} f^{-a}(X_j)} \tag{1.2.9}$$

特别地，称 $a = 1$ 的删除算子 $T_a^1$ 为比例删除算子；称 $\alpha = 0$ 的删除算子为均匀删除算子。

**定理 1.2.7**　(1) 设 $\vec{X}$ 为一个种群，且对于某个 $i \leqslant N$，存在 $j \leqslant N$，使 $f(X_i) > f(X_j)$，则

$$\lim_{a\to\infty} P\{X_i \in T_d^a(\vec{X})\} = 1$$

(2) 对于任意种群 $\vec{X} = (X_1, X_2, \cdots, X_N)$，必有

$$\lim_{a\to 0} P\{X_i \notin T_d^a(\vec{X})\} = \frac{1}{N}$$

**证明**　(1) 由于 $f(X_i) > f(X_j)$，则

$$\begin{aligned}
&\lim_{a\to\infty} P\{X_i \in T_d^a(\vec{X})\} \\
&= 1 - \lim_{a\to\infty} P\{T_d^a(\vec{X}) = \vec{X} \setminus \{X_i\}\} \\
&= 1 - \lim_{a\to\infty}\left(1 + \sum_{k\neq i}\left(\frac{f(X_i)}{f(X_k)}\right)^a\right)^{-1}
\end{aligned}$$

$$\geqslant 1 - \lim_{\alpha \to \infty}\left(1 + \left(\frac{f(X_i)}{f(X_j)}\right)^{\alpha}\right)^{-1} = 1$$

则证(1)。

(2) 对于 $\vec{X} = (X_1, \cdots, X_N)$，有

$$\lim_{\alpha \to 0} P\{X_i \notin T_d^{\alpha}(\vec{X})\}$$

$$= \lim_{\alpha \to 0} P\{T_d^{\alpha}(\vec{X}) = \vec{X} \setminus \{X_i\}\} = \lim_{\alpha \to 0} \frac{f^{-\alpha}(X_i)}{\sum_{k=1}^{N} f^{-\alpha}(X_k)} = \frac{1}{N}$$

随机删除算子中，对于 $\alpha$ 的选择很重要。$\alpha$ 越大时适应值大的个体越容易被保留。$\alpha$ 越小时删除算子越具有随机性，与适应值大小的关系越不重要。删除算子体现了"劣汰"的性质。

有了以上概念，我们可以介绍几种常见的遗传算法。

**标准遗传算法**(canonical genetic algorithm)：

(1) 置 $k = 0$，随机产生初始种群

$$\vec{X}(0) = (X_1(0), \cdots, X_N(0)) \in S^N$$

(2) 独立地从当前种群中选取 $N$ 对母体；

(3) 独立地对于 $N$ 个母体进行杂交得到 $N$ 个中间个体；

(4) 独立地对 $N$ 个杂交后的个体进行变异，得到下代种群

$$\vec{X}(k+1) = (X_1(k+1), \cdots, X_N(k+1)) \in S^N$$

(5) 检验停止准则。若满足则停止，否则 $k := k + 1$ 并返回到(2)。

**杰出选择遗传算法**(elitist selection genetic algorithm)

(1) 置 $k = 0$，随机产生初始种群

$$\vec{X}(0) = (X_1(0), \cdots, X_N(0)) \in S^N$$

(2) 独立地从当前种群中选取 $N - 1$ 个母体；

(3) 独立地对所选 $N - 1$ 个母体进行杂交得到 $N - 1$ 个中间个体；

(4) 独立地对杂交后的 $N - 1$ 个个体进行变异得到 $k + 1$ 代种

群的前 $N-1$ 个个体

$$X_1(k+1),\cdots,X_{N-1}(k+1)$$

(5) 计算 $i=\arg\max_j\{f(X_j(k))\}$，令

$$X_N(k+1)=X_i$$

(6) 检验停止准则。若满足则停止，否则 $k:=k+1$ 并返回(2)

在杰出选择遗传算法中，$i=\arg\max_j\{f(X_j(k))\}$ 表示使 $f(X_j(k))$ 取最大值的个体为 $X_i(k)$。

**稳定状态遗传算法**(steady state genetic algorithm)

(1) 置 $k=0$，随机产生初始种群

$$\vec{X}(0)=(X_1(0),\cdots,X_N(0))\in S^N$$

(2) 从当前种群中选择一个母体；

(3) 对所选母体进行杂交和变异得到一个新个体；

(4) 删去 $\vec{X}(k)$ 中的一个个体而代之以产生的新个体形成下代种群

$$\vec{X}(k+1)=(X_1(k+1),\cdots,X_N(k+1))\in S^N$$

(5) 检验停止准则。若满足则停止，否则 $k:=k+1$ 并返回(2)。

以上是几种常见的遗传算法。变化选择、杂交、变异、删除的方式，还可以得到不同的遗传算法。遗传算法通过遗传机制的“优胜”或“劣汰”，保留好的性质，淘汰差的性质，使遗传过程不断进化。

## 1-3　遗传机制的几何表示

对于一个链长为 $l$ 的个体全体，即个体空间

$$S=\{0,1\}^l$$

可以视为 $l$ 维立方体中的顶点集。例如 $l=3$ 时为3维立方体。

在3维立方体中的一个顶点构成一个个体。对于3维立方体，

只有 12 条直线，每个直线上有两个顶点，两个顶点仅在一个基因位置上取值不同。如(0 1 0) 和(1 1 0) 在一条直线上，可以表示这条直线为( * 1 0)，即第一个基因位置可取0或1，而后两个基因位置保持不变。同样(0 1 0) 和(0 1 1) 也在一条直线上，可以表示这条直线为(0 1 * )，即第三个基因位置可取0或1，而前两个基因位置保持不变。对于 3 维立方体只有 6 个平面，每个平面上有 4 个顶点，4 个顶点在两个基因位置上取值不同。如(0 1 0)，(1 1 0)，(1 1 1)，(0 1 1) 在一个平面上，第二个基因位置保持不变，其它两个基因位置可取0或1。在3维立方体中，直线和平面都称为3维立方体的"超平面"，它们是3维立方体的"子空间"。"超平面"保留某些基因位置不变化，减小搜索范围。我们一旦确定某些基因位置取值是有效的，则可以固定这些基因位置取值，使搜索范围减小，而降低计算工作量。

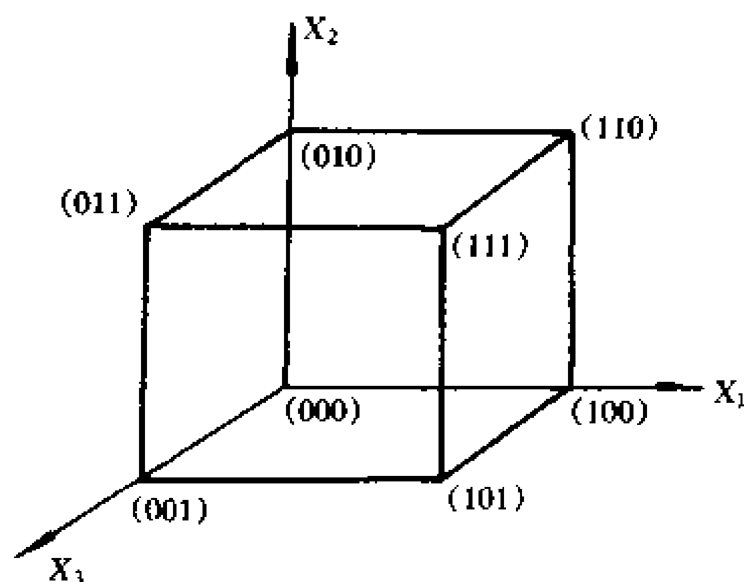



图 1.3.1 $l = 3$ 的个体空间

Holland 在 1975 年引入模式(Schema) 的概念，在遗传算法的理论分析中有着重要意义。

**定义 1.3.1** 设 $a_{i_k} \in \{0,1\}, 1 \leqslant i_k < i_{k+1} \leqslant l$，一个模式是

个体空间中的"超平面",记作

$$\mathscr{L}=[(i_k,a_{i_k});0\leqslant k\leqslant K]$$
$$=\{X=(x_1,\cdots,x_l)\in S;\ x_{i_k}=a_{i_k}(k\leqslant K)\}$$

其中 $K$ 称作模式的阶,记作 $O(\mathscr{L})=K$。$\{i_1,\cdots,i_K\}$ 称 $\mathscr{L}$ 的定义分量(或定义基因位置),$a_{i_k}$ 称作定义分量值(定义等位基因),模式的定义长度定义为

$$\delta(\mathscr{L})=i_K-i_1$$

在不致引起混淆的情况下,将模式 $\mathscr{L}[(i_k,a_{i_k});k\leqslant K]$ 简记作 $\mathscr{L}$ 或 $\mathscr{L}[(i_k,a_{i_k});K]$。

**例 1.3.1**(续例 1.2.1)　取 $a_1=1$,则

$$\mathscr{L}_1[(1,1)]=\{(1\ 0\ 0),(1\ 0\ 1),(1\ 1\ 0),(1\ 1\ 1)\}$$

取 $a_1=1,a_2=0$,则

$$\mathscr{L}_2[(1,1),(2,0)]=\{(1\ 0\ 0),(1\ 0\ 1)\}$$

在例 1.2.1 中 $X_1$ 和 $X_4$ 在一个"超平面"上,即

$$\mathscr{L}_3[(1,0),(2,1)]=\{(0\ 1\ 1),(0\ 1\ 0)\}$$

$X_3$ 和 $X_4$ 在一个"超平面"上,即

$$\mathscr{L}_4[(2,1),(3,0)]=\{(0\ 1\ 0),(1\ 1\ 0)\}$$

但是 $X_1,X_2,X_3,X_4$ 不在任何一个"超平面"上。在上面的例子中,$O(\mathscr{L}_2)=O(\mathscr{L}_3)=O(\mathscr{L}_4)=2,O(\mathscr{L}_1)=1$,而 $\delta(\mathscr{L}_1)=0$,$\delta(\mathscr{L}_2)=1,\delta(\mathscr{L}_3)=1,\delta(\mathscr{L}_4)=1$。

下面我们用模糊度描述模式的阶。

**定义 1.3.2**　称映射 $H:[0,1]\to[0,1]$ 为模糊度,若满足:

(1) $H(a)=0$ 当且仅当 $a=0,1,H(1/2)=1/2$;

(2) $H(a)=H(1-a)$;

(3) $H(a)$ 在$[0,1/2]$上单调增。

**例 1.3.2**　下列映射是模糊度。

$$H_1(a) = 2a(1-a)$$

$$H_2(a) = -a\log_2 a$$

$$H_3(a) = \begin{cases} a, & a \leqslant 1/2 \\ 1-a, & a > 1/2 \end{cases}$$

**定义 1.3.3** 设 $\vec{X} = (X_1, \cdots, X_N)$ 为种群，$X_i = (x_{i1}, \cdots, x_{il})$，$(i \leqslant N)$，$H$ 为模糊度，称

$$d_j(\vec{X}) = H\left(\frac{1}{N}\sum_{i=1}^{N} x_{ij}\right)$$

为 $\vec{X}$ 在分量 $j$ 处的多样度，称

$$d(\vec{X}) = \frac{1}{l}\sum_{j=1}^{l} d_j(\vec{X})$$

为 $\vec{X}$ 的种群多样度。记

$$\lambda(\vec{X}) = \sum_{j=1}^{l} \chi_{\{0\}}(d_j(\vec{X}))$$

$\vec{X}$ 的成熟度为

$$\beta(\vec{X}) = l - \lambda(\vec{X})$$

**例 1.3.3** 设 $S = \{0,1\}^3$，取 $\vec{X} = (X_1, X_2, X_3, X_4)$，且

$$X_1 = (0\ 1\ 1) \qquad X_2 = (0\ 0\ 1)$$

$$X_3 = (1\ 1\ 0) \qquad X_4 = (0\ 1\ 0)$$

且 $H(a) = 2a(1-a)$，则：

$$d_1(\vec{X}) = H\left(\frac{1}{4}\right) = \frac{3}{8}$$

$$d_2(\vec{X}) = H\left(\frac{3}{4}\right) = \frac{3}{8}$$

$$d_3(\vec{X}) = H\left(\frac{1}{2}\right) = \frac{1}{2}$$

$$d(\vec{X}) = \frac{1}{3}\left(\frac{3}{8} + \frac{3}{8} + \frac{1}{2}\right) = \frac{1}{3} \times \frac{10}{8} = \frac{5}{12}$$

$$\lambda(\vec{X}) = 0 \qquad \beta(\vec{X}) = 3$$

**定理 1.3.1**　设 $\vec{X}$ 为一种群，$\mathscr{H} = \mathscr{H}[\vec{X}]$ 为包含 $\vec{X}$ 的最小模式，则 $O(\mathscr{H}) = \lambda(\vec{X}) = l - \beta(\vec{X})$。

**证明**　由于 $d_j(\vec{X}) = 0$，当且仅当 $x_{ij} \equiv 0(i \leqslant N)$ 或 $x_{ij} \equiv 1(i \leqslant N)$，即第 $j$ 个基因位置相同。$\lambda(\vec{X})$ 是 $d_j(\vec{X}) = 0$ 的个数，于是 $\lambda(\vec{X}) = O(\mathscr{H})$。

**定理 1.3.2**　设 $\vec{X}$ 和 $\vec{Y}$ 为两个种群，且 $\vec{X} \sim \vec{Y}$，则

$$d_j(\vec{X}) = d_j(\vec{Y}) \qquad (j \leqslant l)$$

**证明**　由于

$$\frac{1}{N}\sum_{i=1}^{N} x_{ij} = \frac{1}{N}\sum_{X \in S}\varphi(\vec{X}, X)x_j$$

$$\frac{1}{N}\sum_{i=1}^{N} y_{ij} = \frac{1}{N}\sum_{X \in S}\varphi(\vec{Y}, X)x_j .$$

其中 $X = (x_1, \cdots, x_l) \in S$ 为个体。根据假设 $\vec{X} \sim \vec{Y}$，则对于任意 $X \in S$ 有 $\varphi(\vec{X}, X) = \varphi(\vec{Y}, X)$，于是

$$d_j(\vec{X}) = H\left(\frac{1}{N}\sum_{i=1}^{N} x_{ij}\right) = H\left(\frac{1}{N}\sum_{i=1}^{N} y_{ij}\right) = d_j(\vec{Y})$$

如果 $d_j(\vec{X}) = d_j(\vec{Y})\ (j \leqslant l)$，称 $\vec{X}$ 与 $\vec{Y}$ 弱等价，记为 $\vec{X} \overset{(w)}{\sim} \vec{Y}$，则"$\overset{(w)}{\sim}$"也构成等价关系，也可以将 $S^N$ 进行等价分类。由定理1.3.2，若 $\vec{X} \sim \vec{Y}$，则 $\vec{X} \overset{(w)}{\sim} \vec{Y}$，反之则不然。例如

$$\vec{X} = \{(1\,0\,1\,1)(1\,0\,1\,1)(0\,1\,0\,0)(0\,0\,1\,1)\}$$

$$\vec{Y} = \{(1\,0\,1\,1)(1\,0\,1\,1)(0\,1\,0\,1)(0\,0\,1\,0)\}$$

有 $d_j(\vec{X}) = d_j(\vec{Y})\ (j \leqslant 4)$，但是 $\vec{X} \sim \vec{Y}$ 不成立。由此可见，按等价关系"$\overset{(w)}{\sim}$"将 $S^N$ 分类比"$\sim$"分类更粗糙。

模糊度是对基因作用的进一步刻画。若在一个种群的某个基因位置上模糊度为 0，该基因位置在杂交中不起作用，因此减小了杂交的搜索空间。模糊度为 0，即是种群中的基因为不变基因；否则为可变基因。

为方便起见，我们可将模式 $\mathscr{S}$ 表示为一个可变向量，在固定 $i_k$ 位置上取值 $a_{i_k}(k \leqslant K)$，其它位置用变元符号 * 表示，如

$$\mathscr{S}_1 = (1, *, *) \qquad \mathscr{S}_2 = (1, 0, *)$$

因此一个模式可以视为在{0,1,*}中取值。于是对于长度 $l$ 的个体产生了 $3^l$ 个模式。

**例 1.3.4**(续例 1.2.3) 由于杂交 $X_3 = (1\,1\,0)$ 和(0 1 1)的结果为 $X_3' = (1\,1\,1)$，$f(X_3') = 7$，于是杂交结果提高了适应值。我们比较以下的结果：

$$X_1 = (0\,1\,1) \qquad f(X_1) = 3$$
$$X_2 = (0\,0\,1) \qquad f(X_2) = 1$$
$$X_3 = (1\,1\,0) \qquad f(X_3) = 6$$
$$X_4 = (0\,1\,0) \qquad f(X_4) = 2$$
$$X_3' = (1\,1\,1) \qquad f(X_3') = 7$$

显然第一个基因位置的取值对于选择好的适应值是重要的。$\mathscr{S}_3[(1,0)]$ 和 $\mathscr{S}_3[(1,1)]$ 将 $S$ 分为两部分，在 $\mathscr{S}[(1,1)]$ 中可以搜索到适应值大的个体。

**定义 1.3.4** 对于第 $i$ 个基因，若

$$f((x_1, \cdots, x_{i-1}, 1, x_{i+1}, \cdots, x_l)) \geqslant f((x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_l)) \tag{1.3.1}$$

对于任意 $x_k = 0,1(k \neq i)$ 成立，或

$$f((x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_l)) \geqslant f((x_1, \cdots, x_{i-1}, 1, x_{i+1}, \cdots, x_l)) \tag{1.3.2}$$

对于任意 $x_k = 0,1(k \neq i)$ 成立，称基因 $i$ 是重要基因，且式(1.3.1)成立时，重要基因取值 1 时为优良选择；当式(1.3.2)成立时，重要基因取值 0 时为优良选择。若基因 $i$ 是重要基因，1 是优良选择，则最优解必在模式 $\mathscr{S}[(i,1)]$ 中。当 0 是优良选择时，最优

解必在模式 $\mathscr{L}[(i,0)]$ 中。

如果存在一个重要基因，搜索空间就减小了一半，即只要搜索 $2^{l-1}$ 个个体。一般地，若有 $k$ 个重要基因，则搜索空间仅有 $2^{l-k}$ 个个体。

**定义 1.3.5** 设 $\mathscr{L}$ 是一个模式，$\vec{X}\in S^N$ 是一个种群。

(1) 称 $\vec{X}$ 包含于 $\mathscr{L}$ 中，记作 $\vec{X}\subset\mathscr{L}$，若对于任意 $i\leqslant N$，$X_i\in\mathscr{L}$，其中 $\vec{X}=(X_1,\cdots,X_N)$；

(2) 称 $\vec{X}$ 与 $\mathscr{L}$ 相交非空，记作 $\vec{X}\cap\mathscr{L}\neq\varnothing$，若存在 $i\leqslant N$，使 $X_i\in\mathscr{L}$；

(3) 称 $\mathscr{L}[\vec{X}]$ 为包含 $\vec{X}$ 最小模式，若对于任意包含 $\vec{X}$ 的模式 $\mathscr{L}$，必有 $\mathscr{L}[\vec{X}]\subset\mathscr{L}$。

**定理 1.3.3** (1) 设 $\mathscr{L}$ 为一个模式，则 $|\mathscr{L}|=2^{l-O(\mathscr{L})}$；

(2) 若 $\mathscr{L}_1[(i_k,a_{i_k});K_1]$ 和 $\mathscr{L}_2[(j_k,b_{j_k});K_2]$ 为两个模式，且满足条件：

$$a_{i_k}=b_{j_k}\qquad(i_k=j_k)$$

则

$$\mathscr{L}=\mathscr{L}_1[(i_k,a_{i_k});K_1]\cap\mathscr{L}_2[(j_k,b_{j_k});K_2]$$

也是一个模式；

(3) 任给 $\vec{X}\in S^N$，$\mathscr{L}[\vec{X}]$ 必存在且唯一。

**证明** (1) 由于 $O(\mathscr{L})=K$ 表示模式的阶，即在个体中基因位置取值不变的数量，于是基因位置变化的数量为 $l-O(\mathscr{L})$，则证。

(2) 取 $I=\{i_k;k\leqslant K_1\}\cup\{j_k;k\leqslant K_2\}$，则

$$\mathscr{L}=\mathscr{L}_1[(q_k,a_k),(q_k=i_k,k\leqslant K_1);$$
$$(q_k,b_k),(q_k=j_k;k\leqslant K_2)\}$$

于是 $\mathscr{L}$ 为模式。

(3) 假设 $\vec{X}$ 中各个个体在每一个分量上的取值均不相同，此

时对于每一个阶 $O(\mathscr{S})$ 大于零的模式 $\mathscr{S}$，必有 $\vec{X} \not\subset \mathscr{S}$。因此，包含 $\vec{X}$ 的唯一的模式为 $\mathscr{S}[\vec{X}] = S$。假设 $\vec{X}$ 中的个体仅在分量 $j_k (K < l)$ 上有相同的值，即

$$x_{ij_k} = a_{j_k} \qquad (k \leqslant K, i \leqslant N)$$

此时 $\mathscr{S}[\vec{X}] = \mathscr{S}[(j_k, a_{j_k}), K]$。事实上，由(2)知 $\mathscr{S}$ 为包含 $\vec{X}$ 的模式全体之交。

**例 1.3.5**　对于 $S = \{0,1\}^3$，若

$$\mathscr{S}_1 = \mathscr{S}[(2,1)], \qquad \mathscr{S}_2 = \mathscr{S}[(1,1)]$$

$$\mathscr{S}_3 = \mathscr{S}[(2,1),(3,0)]$$

则 $|\mathscr{S}_1| = |\mathscr{S}_2| = 2^{3-1} = 4$，$|\mathscr{S}_3| = 2^{3-2} = 2$，且

$$\mathscr{S}_1 \cap \mathscr{S}_2 = \mathscr{S}[(1,1),(2,1)] = \{(1\ 1\ 1),(1\ 1\ 0)\}$$

$$\mathscr{S}_1 \cap \mathscr{S}_3 = \mathscr{S}_3 = \{(0\ 1\ 0),(1\ 1\ 0)\}$$

由例 1.3.5 可见，$\mathscr{S}_1$ 与 $\mathscr{S}_2$ 仅有三种情形，$\mathscr{S}_1 \cap \mathscr{S}_2 = \varnothing$；$\mathscr{S}_1 \subset \mathscr{S}_2$ 或 $\mathscr{S}_2 \subset \mathscr{S}_1$；$\mathscr{S}_1 \cap \mathscr{S}_2 \neq \varnothing$ 且 $\mathscr{S}_1 \not\subset \mathscr{S}_2$，$\mathscr{S}_2 \not\subset \mathscr{S}_1$。如果两个模式确定了两个重要基因，则它们的交运算产生的模式即是由两个重要基因产生的模式，搜索空间跟着缩小。

对于例 1.2.3 中的初始种群 $\vec{X} = (X_1, X_2, X_3, X_4)$：

$$X_1 = (0\ 1\ 1) \qquad X_2 = (0\ 0\ 1)$$

$$X_3 = (1\ 1\ 0) \qquad X_4 = (0\ 1\ 0)$$

不在一个平面上，因此 $\mathscr{S}[\vec{X}] = S = \{0,1\}^3$。

**定理 1.3.4**　设 $T_s$ 为选择算子，$T_c$ 为杂交算子，而

$$\mathscr{S} = \mathscr{S}[(i_k, a_{i_k}), K]$$

为一个模式，$\vec{X} \subset \mathscr{S}$，则有

$$P\{T_c \circ T_s^n(\vec{X}) \subset \mathscr{S}\} = 1$$

即

$$T_c \circ T_s^n(\vec{X}) \subset \mathscr{S}$$

特别地有

$$T_c \circ T_s^a(\vec{X}) \subset \mathscr{L}[\vec{X}]$$

其中 $\mathscr{L}[\vec{X}]$ 为由 $\vec{X}$ 生成的最小模式。

**证明**　对于任意$(X,Y) \subset \mathscr{L}, Z \notin \mathscr{L}$，则$(X,Y)$杂交后生成 $Z$ 的基因位置个数$k = k(X,Y,Z) = 0$，由定理1.2.3有

$$P\{T_c(X,Y) = Z\} = 0$$

由概率性质及全概率公式有

$$\begin{aligned} & P\{T_c(T_s^a(\vec{X})) \subset \mathscr{L}\} \\ & = 1 - \sum_{Z \notin \mathscr{L}} P\{T_c(T_s^a(\vec{X})) = Z\} \\ & = 1 - \sum_{Z \notin \mathscr{L}} \sum_{(X,Y) \subset \mathscr{L}} P\{T_c(X,Y) = Z\} \\ & \quad \cdot P\{T_s^a(\vec{X})) = (X,Y)\} \\ & = 1 \end{aligned}$$

则证。

定理1.2.1和定理1.3.4表明，任何种群 $\vec{X}$，选择在 $\vec{X}$ 中进行，而杂交在 $\mathscr{L}[\vec{X}]$ 中进行，杂交比选择扩大了搜索范围。选择只在种群的 $N$ 个个体中进行，而杂交后产生的个体是 $2^{l-K}$ 个个体中之一，其中 $K = O(\mathscr{L}[\vec{X}])$。

**例1.3.6**　对于 $S = \{0,1\}^3$，研究初始种群

$$\vec{X} = (X_1, X_2, X_3)$$

其中

$$X_1 = (0\ 1\ 1) \qquad X_2 = (1\ 1\ 0) \qquad X_3 = (0\ 1\ 0)$$

显然

$$\begin{aligned} \mathscr{L}[\vec{X}] &= \mathscr{L}[(2,1)] \\ &= \{(0\ 1\ 0),(0\ 1\ 1),(1\ 1\ 0),(1\ 1\ 1)\} \end{aligned}$$

$X_1$ 与 $X_2$ 在杂交点2处杂交得到：

$$X_1' = (0\ 1\ 0) = X_3 \qquad X_2' = (1\ 1\ 1)$$

$X_1$ 与 $X_2$ 在杂交点 1 处杂交得到：

$$X_1' = (0\ 1\ 0) = X_3 \qquad X_2' = (1\ 1\ 1)$$

不管怎样杂交，第二个基因位置总是 1，于是杂交结果在 $\mathscr{L}[\vec{X}]$ 中，同时也看到，选择只能取 $X_1, X_2$ 和 $X_3$，而杂交产生(1 1 1)不在种群中，可见杂交较选择有更大的个体搜索范围。

我们用 $\mathscr{L}[\vec{X}]$ 表示包含 $\vec{X}$ 的最小种群。记

$$C(\vec{X},0) = \vec{X}$$

$$\begin{aligned} C(\vec{X},k) &= T_c \circ T_s(C(\vec{X},k-1)) \\ &= \{Y;\text{存在 } Y_1, Y_2 \in C(\vec{X},k-1) \\ &\quad \text{使 } T_c(Y_1,Y_2) = Y\}\ (k \geqslant 1) \end{aligned}$$

$$\mathscr{H}(\vec{X}) = \bigcup_{k=0}^{\infty} C(\vec{X},k)$$

**定理 1.3.5**　设 $\vec{X}$ 为一种群，则

$$\mathscr{L}[\vec{X}] = \mathscr{H}(\vec{X})$$

**证明**　由于 $\vec{X} \subset \mathscr{L}[\vec{X}]$，于是 $C(\vec{X},0) = \vec{X} \subset \mathscr{L}[\vec{X}]$。若 $C(\vec{X},k-1) \subset \mathscr{L}[\vec{X}]$，由定理 1.3.2 知

$$C(\vec{X},k) = T_c \circ T_s^n(C(\vec{X},k-1)) \subset \mathscr{L}[\vec{X}]$$

于是由归纳证明即得，对于任意 $k \geqslant 0$ 有

$$C(\vec{X},k) \subset \mathscr{L}[\vec{X}]$$

从而

$$\mathscr{H}[\vec{X}] \subset \mathscr{L}(\vec{X})$$

对于 $Y \in \mathscr{L}(\vec{X})$，必有 $X_i$ 使 $x_{ij} = y_j\ (j \leqslant l)$，于是由均匀随机杂交算子反复进行即得 $Y$。同样用单点杂交算子反复进行也可得 $Y$。于是

$$\mathscr{H}[\vec{X}] \supset \mathscr{L}(\vec{X})$$

则证。

由定理 1.3.5 可知，若 $\vec{X} \cap \mathscr{L} = \varnothing$，则 $P\{T_c \circ T_s^n \subset \mathscr{L}\} = 0$。但是有下面定理。

**定理 1.3.6**　设 $T_s^w$ 为选择算子，$T_c$ 为杂交算子，$\mathscr{S}$ 为模式，$\vec{X}\cap\mathscr{S}\neq\varnothing$，则

$$P\{T_c\circ T_s^w(\vec{X})\subset\mathscr{S}\}>0$$

**证明**　由于 $\vec{X}\cap\mathscr{S}\neq\varnothing$，则不妨设 $X_i\in\mathscr{S}$，由定理 1.2.4

$$P\{T_c\circ T_s^w(\vec{X})=X_i\}>0$$

于是

$$P\{T_c\circ T_s^w(\vec{X})\subset\mathscr{S}\}>0$$

对于变异算子，根据定理 1.2.5 有

$$P\{T_m(X)=Y\}=p_m^{d(X,Y)}(1-p_m)^{l-d(X,Y)}$$

一般来说 $p_m$ 取值较小，于是 $d(X,Y)$ 越小，$P\{T_m(X)=Y\}$ 越大，从而变异算子使个体在它的邻域变化。

根据模式的几何表示可以看到，选择是在种群中进行，选择的结果仍在种群中；杂交是在包含种群 $\vec{X}$ 的最小模式 $\mathscr{S}[\vec{X}]$ 中进行，杂交的结果仍在 $\mathscr{S}[\vec{X}]$ 中；变异是在个体邻域进行，变异的结果在个体 $X$ 的邻域。如果初始种群 $\vec{X}$，$\mathscr{S}[\vec{X}]$ 的阶数 $K$ 很大，则选择、杂交的结果范围很小，仅用选择和杂交不可能越过 $\mathscr{S}[\vec{X}]$。变异虽然在一个个体的邻域进行，它可能越过 $\mathscr{S}[\vec{X}]$。因此，选择和杂交仅在一个“子空间”中搜索，而变异在不断地改变子空间，从而扩大搜索范围，原则上可以扩大到整个子空间。选择、杂交和变异三者结合构成遗传机制。

关于模式与遗传机制的关系，Holland 曾经给出了模式定理：**具有短的定义长度、低阶、并且模式采样的平均适应值在种群平均适应值以上的模式，在遗传算法迭代过程中将按指数增长率被采样。**

我们用 $\mathscr{S}$ 表示固定模式，$O(\mathscr{S})$ 表示模式的阶，$\delta(\mathscr{S})$ 表示模式 $\mathscr{S}$ 的定义长度。$\{\vec{X}(n)\}$ 表示遗传算法序列，$p_c$ 是杂交概率，$p_m$ 为变异概率，$\bar{f}_n$ 表示 $\vec{X}(n)$ 的平均适应值，$\bar{f}_n(\mathscr{S})$ 表示$(\mathscr{S}\cap\vec{X}(n))$

的平均适应值，用 $N(\mathscr{L} \cap \vec{X}(n))$ 表示 $\mathscr{L} \cap \vec{X}(n)$ 中的个体数目，$m(\mathscr{L},n) = E(N(\mathscr{L} \cap \vec{X}(n)))$ 表示 $\mathscr{L}$ 中包含种群 $\vec{X}(n)$ 的个体数目的期望值，则有

$$m(\mathscr{L},n+1)$$
$$\geqslant m(\mathscr{L},n)\frac{\bar{f}(\mathscr{L})}{f_n}\cdot\left[1-p_c\frac{\delta(H)}{l-1}-O(\mathscr{L})p_m\right]$$

由此可见，一个种群若最好的个体包含在 $\mathscr{L}$ 中，则在遗传算法过程中，种群将有更多的个体包含在模式 $\mathscr{L}$ 中，这是遗传机制的吸收性。另一方面变异与杂交不断地破坏这种吸收性，使种群转移到新的模式。

## 1-4　杂交算子的几何性质

选择算子在种群中变化，杂交算子在包含种群的最小模式中变化，变异算子在个体的邻域变化。我们现在进一步讨论杂交算子的几何性质，即杂交算子在模式中的变化形式。

**定义 1.4.1**　设 $S$ 为个体空间，$U = S$ 为指标集，对于任意 $u \in U$，给出映射：

$$\varphi_u : S \to S \qquad \bar{\varphi}_u : S \to S$$
$$h : S^2 \to S$$

若满足条件：对于任意 $X,Y,Z \in S$ 及 $u \in U$ 有

$$h(\varphi_u(X),\bar{\varphi}_u(Y)) = Z \Leftrightarrow \varphi_u(X) = \varphi_u(Z),\bar{\varphi}_u(Y) = \bar{\varphi}_u(Z) \tag{1.4.1}$$

称

$$C_u(X,Y) = h(\varphi_u(X),\bar{\varphi}_u(Y))$$
$$\bar{C}_u(X,Y) = h(\varphi_u(Y),\bar{\varphi}_u(X)) \tag{1.4.2}$$

为母体 $(X,Y) \in S^2$ 的杂交后代，记为

$$T_u(X,Y)=(C_u(X,Y),\bar{C}_u(X,Y))$$

**例1.4.1**　设 $S=\{0,1\}^l$ 为个体空间，对于 $X,Y\in S$，

$$X=(x_1,x_2,\cdots,x_l)$$

$$Y=(y_1,y_2,\cdots,y_l)$$

记

$$X\oplus Y=(x_1\oplus y_1,x_2\oplus y_2,\cdots,x_l\oplus y_l)$$

$$X\otimes Y=(x_1\otimes y_1,x_2\otimes y_2,\cdots,x_l\otimes y_l)$$

$$\bar{X}=(1\oplus x_1,1\oplus x_2,\cdots,1\oplus x_l)=\mathbf{1}\oplus x$$

其中

$$x_i\oplus y_i=\begin{cases}1, & x_i\neq y_i\\ 0, & x_i=y_i\end{cases}$$

$$x_i\otimes y_i=\begin{cases}1, & x_i=y_i=1\\ 0, & \text{否则}\end{cases}$$

令 $U=S$，对于 $u\in S$，记：

$$\varphi_u(X)=u\otimes X,\ \bar{\varphi}_u(X)=(\mathbf{1}\oplus u)\otimes X$$

$$h(X,Y)=X\oplus Y$$

其中 $\mathbf{1}=(1,1,\cdots,1)\in S$ 表示各分量恒为1的向量。由于

$$u\otimes(\mathbf{1}\oplus u)=\mathbf{0}=(0,0,\cdots,0)\in S$$

则 $u$ 与 $(\mathbf{1}\oplus u)$ 是正交的，于是由

$$(u\otimes Z)\oplus(\bar{u}\otimes Z)=Z=\varphi_u(X)\oplus\bar{\varphi}_u(Y)$$

即得

$$\varphi_u(Z)\oplus\bar{\varphi}_u(Z)=\varphi_u(X)\oplus\bar{\varphi}_u(Y)$$

则

$$\varphi_u(X)=\varphi_u(Z)\qquad\bar{\varphi}_u(Y)=\bar{\varphi}_u(Z)$$

即式(1.4.1)成立。因此

$$C_u(X,Y)=(u\otimes X)\oplus(\bar{u}\otimes Y)$$

$$\bar{C}_u(X,Y)=(u\otimes Y)\oplus(\bar{u}\otimes X)$$

为两个个体 $X$ 与 $Y$ 的后代。

从例1.4.1可以进一步分析杂交运算的几何意义。$S = \{0,1\}^l$ 可以视为 $l$ 维空间，$u$ 与 $\bar{u} = 1 \oplus u$ 为 $S$ 中的两个正交向量。$u$ 与 $\bar{u}$ 构成了 $S$ 中的“超平面”。$\varphi_u(X)$ 为 $X$ 在 $u$ 上的投影向量，$\bar{\varphi}_u(Y)$ 为 $Y$ 在 $\bar{u}$ 的投影向量。$C_u(X,Y)$ 即是 $\varphi_u(X)$ 与 $\bar{\varphi}_u(Y)$ 之和，$\bar{C}_u(X,Y)$ 是 $\varphi_u(Y)$ 与 $\bar{\varphi}_u(X)$ 之和。

$u$ 是一个杂交向量。在例1.4.1中，当 $u_i = 1$ 时保留 $X$ 中的 $x_i$；当 $u_i = 0$ 时代替 $Y$ 中的 $y_i$ 而得到后代 $C_u(X,Y)$。同理，当 $u_i = 1$ 时保留 $Y$ 中的 $y_i$；当 $u_i = 0$ 时，代替 $X$ 中的 $x_i$ 而得到 $\bar{C}_u(X,Y)$。当 $x_i = y_i$ 时，子代 $C_u(X,Y)$ 与 $\bar{C}_u(X,Y)$ 中的第 $i$ 个基因通过 $u$ 杂交不会变化。于是 $C_u(X_i,Y_i) \in \mathscr{Q}[\vec{X}]$，$\bar{C}_u(X_i,Y_i) \in \mathscr{Q}[\vec{X}]$。图1.4.1指出，当 $u$ 固定时，$C_u(X,Y)$ 与 $\bar{C}_u(X,Y)$ 是 $X$ 与 $Y$ 的平移。若 $u$ 变化，而 $(X,Y)$ 不变化，则 $C_u(X,Y)$ 与 $\bar{C}_u(X,Y)$ 是 $X$ 与 $Y$ 的旋转。

图1.4.1 杂交运算的几何表示

**定理1.4.1** 对于任意 $(X,Y) \in S^2$，$u \in U$ 以下性质成立：

(1) $C_u(X,X) = \bar{C}_u(X,X) = X$

(2) $C_u(C_u(X,Y),\bar{C}_u(X,Y)) = X$

$$\bar{C}_u(C_u(X,Y),\bar{C}_u(X,Y)) = Y$$

(3) $T_u(T_u(X,Y)) = (X,Y), T_u^{-1} = T_u$

(4) 对于任意 $M \subset S^2$,有

$$T_u(M \cup T_u(M)) = M \cup T_u(M)$$

(5) $\tau_u(M) = M \cup T_u(M)$ 具有以下性质:

$$M \subset \tau_u(M)$$

$$M \subset M' \Rightarrow \tau_u(M) \subset \tau_u(M')$$

$$\tau_u(\tau_u(M)) = \tau_u(M)$$

**证明**　(1) 由于 $\varphi_u(X) = \varphi_u(X), \bar{\varphi}_u(X) = \bar{\varphi}_u(X)$, 由式(1.4.1)则证 $h(\varphi_u(X),\bar{\varphi}_u(X)) = X$。

(2) 设

$$Z = C_u(X,Y),\ W = \bar{C}_u(X,Y)$$

由式(1.4.1)知

$$\varphi_u(Z) = \varphi_u(X) \qquad \bar{\varphi}_u(Z) = \bar{\varphi}_u(Y)$$

$$\varphi_u(W) = \varphi_u(Y) \qquad \bar{\varphi}_u(W) = \bar{\varphi}_u(X)$$

由(1)证

$$X = h(\varphi_u(X),\bar{\varphi}_u(X)) = h(\varphi_u(Z),\bar{\varphi}_u(W))$$
$$= C_u(Z,W) = C_u(C_u(X,Y),\bar{C}_u(X,Y))$$

类似可证

$$Y = \bar{C}_u(C_u(X,Y),\bar{C}_u(X,Y))$$

(3) 由(2)即得。

(4) 若$(X,Y) \in M$,则 $T_u(X,Y) \in T_u(M)$,而 $T_u(T_u(X,Y)) = (X,Y) \in M$,于是

$$T_u(M \cup T_u(M)) \subset M \cup T_u(M)$$

若$(X,Y) \in M$, 则 $T_u(T_u(X,Y)) = (X,Y)$, 则$(X,Y) \in T_u(T_u(M))$,于是

$$M \cup T_u(M) \subset T_u(M \cup T_u(M))$$

则证。

(5) 由(4) 则证。

**定理 1.4.2** 设 $X,Y \in S, u \in U$, $\|\cdot\|$ 为 $S$ 上的 Hamming 范数, $C_u(X,Y) = \varphi_u(X) \oplus \bar{\varphi}_u(Y)$, $\bar{C}_u(X,Y) = \varphi_u(Y) \oplus \bar{\varphi}_u(X)$, 则有

(1) $X \oplus Y = C_u(X,Y) \oplus \bar{C}_u(X,Y)$

(2) $\|X \ominus Y\| = \|C_u(X,Y) \ominus \bar{C}_u(X,Y)\|$

其中 $\ominus$ 等同于 $\oplus$。

**证明** (1) 由于

$$C_u(X,Y) = \varphi_u(X) \oplus \bar{\varphi}_u(Y)$$
$$\bar{C}_u(X,Y) = \varphi_u(Y) \oplus \bar{\varphi}_u(X)$$

于是由定理 1.4.1 即证

$$\begin{aligned} & C_u(X,Y) \oplus \bar{C}_u(X,Y) \\ & = \varphi_u(X) \oplus \bar{\varphi}_u(Y) \oplus \varphi_u(Y) \oplus \bar{\varphi}_u(X) \\ & = C_u(X,X) \oplus \bar{C}_u(Y,Y) = X \oplus Y \end{aligned}$$

(2) 由于

$$\|X \oplus Y\| = \|X \ominus Y\|$$

由(1) 则证。

定理 1.4.1 和定理 1.4.2 指出了杂交运算的几何性质。自身杂交到自身;一个母体杂交后的子代母体再杂交还原为父代;一组种群及其杂交的后代再经过杂交不再变化;杂交后子代之间距离与父代之间距离保持不变;杂交 $T_u$ 是保距变换。

**定义 1.4.2** 对于任意 $X \in S, u \in U$,称

$$\Phi_u(X) = \varphi_u^{-1} \circ \varphi_u(X) = \{Z; \varphi_u(Z) = \varphi_u(X)\}$$

为 $u$ 与 $X$ 确定的模式。

**例 1.4.2** 设 $u = (1\ 0\ 1\ 1\ 0)$, $X = (1\ 0\ 1\ 0\ 1)$, 则

$$\varphi_u(X) = u \otimes X = (1\ 0\ 1\ 0\ 0)$$

于是

$$\Phi_u(X) = \{(1\ 0\ 1\ 0\ 0),(1\ 0\ 1\ 0\ 1),(1\ 1\ 1\ 0\ 0),(1\ 1\ 1\ 0\ 1)\}$$
$$= \mathscr{L}[(1,1),(3,1),(4,0)]$$
$$= (1\ *\ 1\ 0\ *)$$

是一个3阶模式。

同样，对于 $X \in S, u \in U$，

$$\bar{\Phi}_u(X) = \bar{\varphi}_u^{-1} \circ \bar{\varphi}_u(X) = \{Z; \bar{\varphi}_u(Z) = \bar{\varphi}_u(X)\}$$

也确定了一个模式。易得

$$\bar{\Phi}_u(X) = \mathscr{L}[(2,0),(5,1)] = (*,0,*,*,1)$$

且 $\Phi_u(X) \cap \bar{\Phi}_u(X) = X$，即仅有一个公共个体。

**定理1.4.3**　对于任意 $u \in U, X \in S$，由

$$\Phi_u(X) = \varphi_u^{-1} \circ \varphi_u(X) = \{Z; \varphi_u(Z) = \varphi_u(X)\}$$

确定的模式为

$$\Phi_u(X) = \mathscr{L}[(i,x_i); u_i = 1]$$

反之，若有模式

$$\mathscr{L} = \mathscr{L}[(i_k, a_{i_k}); k \leqslant K]$$

则存在 $u \in U, X \in S$，使

$$\mathscr{L} = \Phi_u(X)$$

**证明**　设 $u \in U, X \in S$，由 $Z \in \Phi_u(X)$，对于

$$Z = (z_1, \cdots, z_l)$$
$$u = (u_1, \cdots, u_l)$$

$z_i \otimes u_i = z_i \otimes x_i$，当且仅当 $u_i = 1$ 时 $z_i = x_i$，于是

$$\Phi_u(X) = \mathscr{L}[(i,x_i); u_i = 1]$$

反之，对于模式

$$\mathscr{L} = \mathscr{L}[(i_k, a_{i_k}); k \leqslant K]$$

令 $u=(u_1,\cdots,u_l)$ 为

$$u_i=\begin{cases}1,\ i=i_k(k\leqslant K)\\0,\ 反之\end{cases}$$

$X=(x_1,\cdots,x_l)$ 为

$$x_i=\begin{cases}a_{i_k},\ i=i_k(k\leqslant K)\\ *,\ 反之\end{cases}$$

易证

$$\Phi_u(X)=\mathscr{L}[(i_k,a_{i_k});k\leqslant K]$$

**定理 1.4.4** 对于任意 $u\in U,X\in S$,由

$$\bar{\Phi}_u(X)=\bar{\varphi}_u^{-1}\bar{\varphi}_u(X)=\{Z;\bar{\varphi}_u(Z)=\bar{\varphi}_u(X)\}$$

确定的模式为

$$\bar{\Phi}_u(X)=\mathscr{L}[(i,x_i);u_i=0]$$

反之,若有模式

$$\mathscr{L}=\mathscr{L}[(i_k,a_{i_k});k\leqslant K]$$

则存在 $u\in U,X\in S$,使

$$\mathscr{L}=\bar{\Phi}_u(X)$$

**证明** 类似定理 1.4.3 可证。

**定理 1.4.5** 设 $u,u'\in U,X,Y\in S$

(1) 若 $\varphi_u\circ\varphi_{u'}=\varphi_u$,则 $C_{u'}(\cdot,Y)$ 映射 $\Phi_u(X)$ 到自身;

(2) 若 $\varphi_u\circ\bar{\varphi}_{u'}=\varphi_u$,则 $\bar{C}_{u'}(\cdot,Y)$ 映射 $\Phi_u(X)$ 到自身。

**证明** 设 $Z\in\Phi_u(X),\varphi_u\circ\varphi_{u'}=\varphi_u$,由(1.4.1)有

$$\begin{aligned}\varphi_u(C_{u'}(Z,Y))&=\varphi_u(h(\varphi_{u'}(Z),\bar{\varphi}_{u'}(Y)))\\&=\varphi_u\circ\varphi_{u'}(h(\varphi_{u'}(Z),\bar{\varphi}_{u'}(Y)))\\&=\varphi_u\circ\varphi_{u'}(Z)=\varphi_u(Z)=\varphi_u(X)\end{aligned}$$

(2) 类似(1)可证。

单点杂交相当于取 $u'=2^k-1(k\leqslant l)$,即 $u'$ 为形式

$$u' = (0,\cdots,0,1,\cdots,1)$$

这时 $\bar{C}_u(X,Y)$ 即是单点杂交的结果。对于两点杂交相当于取

$$u' = (0,\cdots,0,1,\cdots,1,0,\cdots,0)$$

这时 $\bar{C}_u(X,Y)$ 即是两点杂交的结果。

**定理 1.4.6**　对于单点杂交 $u' = 2^k - 1(k \leqslant l)$，$\varphi_u \circ \varphi_{u'} = \varphi_u$，当且仅当 $u \leqslant u'$，即 $u'$ 中最左边的 1 落在模式 $\Phi_u(X)$ 固定基因位置的左边。

**证明**　设

$$u' = (0,\cdots,0,1,\cdots,1)$$

$$u = (u_1,\cdots,u_l)$$

$u \leqslant u'$ 当且仅当 $u_i \leqslant u_i'(i \leqslant l)$。设 $u'$ 中的 1 最左边位置为 $k_0 = l - k + 1$，则 $u_i = 0(i \leqslant k_0 - 1)$，于是对于任意

$$X = (x_1,\cdots,x_l)$$

$$\varphi_{u'}(X) = (0,\cdots,0,x_{k_0},\cdots,x_l)$$

$$\varphi_u \circ \varphi_{u'}(X) = \varphi_u(X)$$

则证。

**定理 1.4.7**　对于单点杂交 $u' = 2^k - 1(k \leqslant l)$，$\varphi_u \circ \bar{\varphi}_{u'} = \varphi_u$，当且仅当 $u \leqslant \bar{u}'$，即 $u'$ 中的杂交点落在模式 $\Phi_u(X)$ 的固定基因位置的右边。

**证明**　类似定理 1.3.7。

下面我们进一步分析杂交算子的集体性质。

设 $S_0 \subset S$ 是一族个体，$F \subset U = S$ 是一族杂交向量。记

$$C'(S_0,F,0) = S_0$$

$$\begin{aligned}C'(S_0,F,k) &= T_c(C'(S_0,F,k-1)) \\ &= \{Z;\text{存在 } X,Y \in C'(S_0,F,k-1) \text{ 及 } u \in F,\text{使} \\ &\quad Z = C_u(X,Y) \text{ 或 } Z = \bar{C}_u(X,Y)\}\end{aligned}$$

$$\mathscr{H}(S_0,F)=\bigcup_{k=0}^{\infty}C'(S_0,F,k) \tag{1.4.3}$$

则 $\mathscr{H}(S_0,F)$ 是通过 $F$ 中杂交算子，对于 $S_0$ 中所有个体递归杂交的全体个体。对于 $A\subset F$，记

$$T_A=\bigotimes_{u_i\in A}u_i\bigotimes_{u_j\in F\setminus A}\bar{u}_j$$

$$\mathscr{A}(F)=\{T_A;A\subset F\}$$

$$ST\{S_0,F\}=\{Z\in S;\text{对于任意 }T\in\mathscr{A}(F),\ \text{存在 }X\in S_0,\text{使得 }T\otimes X=T\otimes Z\}$$

**定理 1.4.8** 设 $S_0\subset S,F\subset U$，则

$$\mathscr{H}(S_0,F)=ST\{S_0,F\} \tag{1.4.4}$$

**证明** 首先易证

$$C'(S_0,F,0)=S_0\subset ST(S_0,F)$$

归纳假设

$$C'(S_0,F,k-1)\subset ST(S_0,F)$$

考虑

$$C'(S_0,F,k)=T_c(C'(S_0,F,k-1))$$

设 $Z\in C'(S_0,F,k)$，则存在 $X,Y\in C'(S_0,F,k-1)$ 及 $u\in F$，使

$$Z=C_u(X,Y)\text{ 或 }Z=\bar{C}_u(X,Y)$$

不妨设 $Z=C_u(X,Y)$，则对任意 $T_A\in\mathscr{A}(F)$ 有

$$\begin{aligned}T_A\otimes Z&=T_A\otimes C_u(X,Y)\\&=T_A\otimes[(u\otimes X)\oplus(\bar{u}\otimes Y)]\\&=(u\otimes T_A\otimes X)\oplus(\bar{u}\otimes T_A\otimes Y)\\&=\begin{cases}T_AX,&u\in A\\T_AY,&u\notin A\end{cases}\end{aligned}$$

由于 $X,Y\in ST(S_0,F)$，则对于任意 $T_A\in\mathscr{A}(F)$，存在 $X',Y'\in$

$S_0$，使 $T_A X = T_A X'$，$T_A Y = T_A Y'$，于是 $T_A \otimes Z = T_A \otimes X'$ 或 $T_A \otimes Z = T_A \otimes Y'$，即 $Z \in ST(S_0, F)$，即归纳证明了；对于任意 $k \geqslant 0$ 有 $C'(S_0, F, k) \subset ST(S_0, F)$，从而

$$\mathscr{H}(S_0, F) \subset ST(S_0, F)$$

下面我们证明

$$ST(S_0, F) \subset \mathscr{H}(S_0, F) \tag{1.4.5}$$

设 $Z \in ST(S_0, F)$，则对于任意 $T_A \in \mathscr{P}(F)$，存在 $X \in S_0$，使 $T_A X = T_A Z$。需证明 $Z \in \mathscr{H}(S_0, F)$。证明使用归纳法。对于 $|F| = 1$，$\mathscr{P}(F) = \{T_0, T_1\}$，$T_0 = u$，$T_1 = \bar{u}$，$u \oplus \bar{u} = 1$。于是

$$Z = u \otimes Z \oplus \bar{u} \otimes Z = T_0 \otimes Z \oplus T_1 \otimes Z$$

由于 $Z \in ST(S_0, F)$，故存在 $X \in S_0$，$Y \in S_0$，使 $T_0 \otimes Z = T_0 \otimes X$，$T_1 \otimes Z = T_1 \otimes Y$，于是

$$\begin{aligned} Z &= (T_0 \otimes X) \oplus (T_1 \otimes Y) = (u \otimes X) \oplus (\bar{u} \otimes Y) \\ &= C_u(X, Y) \in \mathscr{H}(S_0, F) \end{aligned}$$

现假定 $|F| = k - 1$ 时式(1.4.5)成立。对于 $|F| = k$ 有

$$F = F_{k-1} \cup \{u\}$$

其中 $|F_{k-1}| = k - 1$。记

$$\mathscr{P}(F) = \{T_A'\}, \qquad \mathscr{P}(F_{k-1}) = \{T_A\}$$

则

$$\begin{aligned} T'_{A \cup \{u\}} &= T'_{A,u} = u \otimes T_A \\ T'_{A \cup \{\bar{u}\}} &= T'_{A,\bar{u}} = \bar{u} \otimes T_A \end{aligned}$$

假设

$$\sum_{A \subset F_{k-1}} T_A = 1$$

则

$$\sum_{A' \subset F} T_{A'}' = \left(u \otimes \sum_{A \subset F_{k-1}} T_A\right) \oplus \left(\bar{u} \otimes \sum_{A \subset F_{k-1}} T_A\right) = 1$$

对于任意 $Z \in ST(S_0,F)$，存在 $X_{A,u}, X_{A,u} \in S_0$ 使得

$$Z = \sum_{A\subset F} T_A Z$$
$$= \left(u \otimes \sum_{A\subset F_{k-1}} T_A Y_{A,u}\right) \oplus \left(\bar{u} \otimes \sum_{A\subset F_{k-1}} T_A Y'_{A,u}\right)$$

由于 $s = t$ 时，$T_s \otimes T_t = T_s$，$s \neq t$ 时，$T_s \otimes T_t = 0$，则

$$\sum_{A\subset F_{k-1}} T_A Y_{A,u},\ \sum_{A\subset F_{k-1}} T_A Y'_{A,u} \in ST(S_0, F_{k-1})$$

从而属于 $\mathscr{H}(S_0,F_{k-1})$，于是 $Z \in \mathscr{H}(S_0,F)$，则证。

对于 $S_0 = \vec{X}$，我们有

$$\mathscr{H}(\vec{X},F) = ST(\vec{X},F) \tag{1.4.6}$$

式(1.4.6) 表达了 $\vec{X}$ 经过 $F$ 中所有杂交向量杂交后的构成，$ST(\vec{X},F)$ 反映了式(1.4.1) 的性质。

# 1-5 遗传机制的过程分析

在遗传算法中，有选择、杂交和变异机制，对于这些机制的性能的分析是遗传算法分析的基础，而这种性能分析必须在遗传过程中进行；为了理论分析需要，我们考虑无限种群的情形。

设 $S = \{0,1\}^l$ 为个体空间，不妨用 $i$ 表示 $S$ 中个体。若 $|S| = M$，则 $i = 0 \sim M-1$。假定 $P$ 为 $S$ 上的概率分布。

**定义 1.5.1** 称 $S$ 上的随机变量 $\hat{X}$ 为无限种群。

记

$$a_i = P\{\hat{X} = i\} \qquad (i = 0 \sim M-1)$$

则

$$A = (a_0, \cdots, a_{M-1}) = (a_i; i \in S)$$

为 $\hat{X}$ 的分布向量。因此，我们也称分布向量 $A$ 为无限种群。

种群是无限种群的特殊情形。对于例 1.2.1 中的种群 $X_1 = 3$，

$X_2=1,X_3=6,X_4=2$上分布为

$$P\{\hat{X}=3\}=P\{\hat{X}=1\}=P\{\hat{X}=6\}=P\{\hat{X}=2\}=\frac{1}{4}$$

$$P\{\hat{X}=0\}=P\{\hat{X}=4\}=P\{\hat{X}=5\}=P\{\hat{X}=7\}=0$$

于是得到

$$A=(0,\ 0.25,\ 0.25,\ 0.25,\ 0,\ 0,\ 0.25,\ 0)$$

由此可见,种群是无限种群的一个样本。

取$\delta_{ij}$为$\delta$函数,$\delta_{ij}=1$当且仅当$i=j$,$\delta_{ij}=0$当且仅当$i\neq j$,记

$$F_{ij}=\delta_{ij}f(i)\qquad(i,j\in S)$$

其中$f(i)$为$i\in S$的适应值。于是得到$M\times M$阶矩阵$\boldsymbol{F}$。

用$\mathbf{1}$表示$M$个分量全为1的向量,依照适应值的选择算子,则有

$$\hat{\boldsymbol{F}}(A)=\frac{\boldsymbol{F}\cdot A}{\mathbf{1}^{T}(\boldsymbol{F}\cdot A)}\tag{1.5.1}$$

也即对于任意$i\in S$有

$$\hat{\boldsymbol{F}}(A)_i=\frac{f(i)\cdot a_i}{\sum_{j\in S}f(j)\cdot a_j}\tag{1.5.2}$$

$\hat{\boldsymbol{F}}(A)$即是在具有分布$A$的无限种群基础上,通过适应值选择而得到的新的无限种群。

式(1.5.1)的过程可以重复进行,即

$$\hat{\boldsymbol{F}}^{(2)}(A)=\hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}(A))$$
$$\vdots$$
$$\hat{\boldsymbol{F}}^{(n)}(A)=\hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}^{(n-1)}(A))$$

这个过程即构成了按照适应值选择的过程。

**例1.5.1**　假定$S=\{0,1\}^2$,有适应值函数:

$$f(0)=6\qquad f(1)=1$$
$$f(2)=2\qquad f(3)=3$$

若给出 $S$ 上的无限种群 $\tilde{X}$ 为:

$$a_0 = P(\tilde{X} = 0) = 0.2 \qquad a_1 = P(\tilde{X} = 1) = 0.4$$
$$a_2 = P(\tilde{X} = 2) = 0.3 \qquad a_3 = P(\tilde{X} = 3) = 0.1$$

由式(1.5.2)即得

$$\hat{F}(A) = (0.48\ 0.16\ 0.24\ 0.12)^{\mathrm{T}}$$

同理可得

$$\hat{F}^{(2)}(A) = \hat{F}(\hat{F}(A)) = (0.75\ 0.041\ 0.164\ 0.086)^{\mathrm{T}}$$

一直进行下去可得

$$\lim_{n\to\infty}\hat{F}(A) = (1\ 0\ 0\ 0)^{\mathrm{T}}$$

于是 $f$ 在 $i = 0$ 时有最大值。

用 $T_s(n)$ 表示经过 $n$ 步适应值选择得到的个体,则

$$\begin{aligned}\hat{F}^{(n)}(A)_i &= P\{T_s(n) = i/A\}\\ &= P\{T_s(1) = i/\hat{F}^{(n-1)}(A)\}\\ &= \hat{F}(\hat{F}^{(n-1)}(A))_i\end{aligned} \tag{1.5.3}$$

**定理 1.5.1** 对于无限种群的适应值选择有

$$\hat{F}^{(n)}(A)_i = \frac{f^n(i)\cdot a_i}{\sum_{j\in S} f^n(j)\cdot a_j} \tag{1.5.4}$$

**证明** 当 $n = 1$ 时即是式(1.5.2)。若式(1.5.4)对指标 $n$ 成立,由于式(1.5.3),则

$$\begin{aligned}\hat{F}^{(n+1)}(A)_i &= \hat{F}(\hat{F}^{(n)}(A))_i\\ &= \frac{f(i)\cdot \dfrac{f^n(i)a_i}{\sum_{j\in S} f^n(j)\cdot a_j}}{\sum_{j\in S} f(j)\dfrac{f^n(j)a_j}{\sum_{j\in S} f^n(j)\cdot a_j}}\\ &= \frac{f^{n+1}(i)\cdot a_i}{\sum_{j\in S} f^{n+1}(j)\cdot a_j}\end{aligned}$$

由归纳法则证。

**定理 1.5.2**　设 $S$ 上的初始无限种群 $\bar{X}$ 有分布

$$A = (a_i;\ i \in S)$$

记

$$S_0 = \{i;\ a_i > 0\}$$

$$S_1 = \{i \in S_0;\ f(i) = \max_{j \in S_0} f(j)\}$$

于是

$$\bar{a}_i = \lim_{n \to \infty} P\{T_s(n) = i/A\} = \begin{cases} \dfrac{a_i}{\sum\limits_{j \in S_1} a_j}, & i \in S_1 \\ 0, & i \notin S_1 \end{cases} \tag{1.5.5}$$

且存在 $C > 0$ 使

$$\sum_{i \in S} | P\{T_s(n) = i/A\} - \bar{a}_i | \leqslant C * \lambda^{-n} \tag{1.5.6}$$

其中

$$\lambda = \min\left\{\frac{f_{\max}}{f(i)};a_i > 0 \text{ 且 } i \notin S_1\right\}$$

$f_{\max}$ 为 $f$ 的最大值,即 $f_{\max} = f(i_0)$,($i_0 \in S_1$)。

**证明**　由定理 1.5.1 即得

$$\lim_{n \to \infty} P\{T_s(n) = i/A\} = \lim_{n \to \infty} \frac{(f(i))^n \cdot a_i}{\sum\limits_{j \in S} (f(j))^n \cdot a_j}$$

$$= \lim_{n \to \infty} \frac{a_i}{\sum\limits_{j \in S} \left(\dfrac{f(j)}{f(i)}\right)^n \cdot a_j}$$

$$= \frac{a_i}{\sum\limits_{j \in S} \lim\limits_{n \to \infty} \left(\dfrac{f(j)}{f(i)}\right)^n \cdot a_j}$$

由于 $i \in S_1$ 时，$\lim\limits_{n\to\infty}\left(\dfrac{f(j)}{f(i)}\right)^n = 0$ 当且仅当 $j \notin S_1$。若 $i \notin S_1$ 时，$\lim\limits_{n\to\infty}\left(\dfrac{f(j)}{f(i)}\right)^n = \infty$ 当且仅当 $f(j) > f(i)$。则证式(1.5.5)。

下面证明式(1.5.6)。由于

$$
\begin{aligned}
&\sum_{i\in S} |P\{T_s(n) = i/A\} - \bar{a}_i| \\
&= \sum_{i\in S_1}\left|\frac{f_{\max}^n \cdot a_i}{\sum\limits_{j\in S}(f(j))^n \cdot a_j} - \frac{a_i}{\sum\limits_{j\in S_1} a_j}\right| \\
&\quad + \sum_{i\notin S_1}\left|\frac{(f(i))^n \cdot a_i}{\sum\limits_{j\in S}(f(j))^n \cdot a_j}\right| \\
&= \sum_{i\in S_1}\left|\frac{a_i}{\sum\limits_{j\in S_1} a_j + \sum\limits_{j\notin S_1}\left(\frac{f(j)}{f_{\max}}\right)^n a_j} - \frac{a_i}{\sum\limits_{j\in S_1} a_j}\right| \\
&\quad + \sum_{i\notin S_1}\left|\frac{a_i}{\sum\limits_{j\in S_1}\left(\frac{f_{\max}}{f(i)}\right)^n a_j + \sum\limits_{j\notin S_1}\left(\frac{f(j)}{f(i)}\right)^n a_j}\right| \\
&\leqslant \sum_{i\in S_1}\left|\frac{\sum\limits_{j\notin S_1}\left(\frac{f(j)}{f_{\max}}\right)^n a_j}{\sum\limits_{j\in S_1} a_j + \sum\limits_{j\notin S_1}\left(\frac{f(j)}{f_{\max}}\right)^n a_j}\right| \frac{a_i}{\sum\limits_{j\in S_1} a_j} \\
&\quad + \sum_{i\notin S_1}\left|\frac{a_i}{\lambda^n \sum\limits_{j\in S_1} a_j}\right| \\
&\leqslant \sum_{i\in S_1}\frac{a_i}{\sum\limits_{j\in S_1} a_j} \cdot \left|\frac{\lambda^{-n}\sum\limits_{j\notin S_1} a_j}{\sum\limits_{j\in S_1} a_j}\right| + \sum_{i\notin S_1}\frac{a_i}{\lambda^n \sum\limits_{j\in S_1} a_j}
\end{aligned}
$$

$$\leqslant \lambda^{-n} \cdot \left\{2 \cdot \frac{\sum_{j \notin S_1} a_j}{\sum_{j \in S_1} a_j}\right\}$$

则证。

定理 1.5.2 指出，对于任意初始的无限种群，按适应值选择算子重复进行，最终可以实现使 $f(i)$ 最大的 $i$ 个体有较高的概率，只要在初始无限种群中存在 $i$ 个体，使 $f(i) = \max_{j \in S} f(j)$。$\lambda$ 的大小决定了收敛速度，$\lambda$ 很大时，收敛速度是很快的。$\lambda$ 很小时，收敛速度比较慢。$\lambda$ 的大小依赖于最大值与次最大值之比，显然有 $\lambda > 1$。对于有限种群，由于选择算子只能在初始种群中进行，不断重复选择算子的结果仅仅可以在初始种群中找到最大值。由此可见，仅仅采取依适应值的选择算子，使遗传算法收敛的必要条件是初始种群包含有全局最优解。

**定理 1.5.3**　设 $S$ 上的初始无限种群 $\tilde{X}$ 有分布

$$a_j = \begin{cases} \dfrac{1}{N}, & j \in S_0, \ |S_0| = N \\ 0, & j \notin S_0 \end{cases}$$

则

$$\sum_{i \in S} |P\{T_s(n) = i/A\} - \bar{a}_i|$$

$$= 2 - 2|S_1| \frac{f_{\max}^n}{\sum_{j \in S_0} f^n(j)} \tag{1.5.7}$$

其中

$$\bar{a}_i = \lim_{n \to \infty} P\{T_s(n) = i/A\}$$

$$= \begin{cases} \dfrac{1}{|S_1|}, & i \in S_1 \\ 0, & i \notin S_1 \end{cases}$$

$S_1 = \{i; f(i) = \max\limits_{j \in S} f(j)\}$。

**证明** 首先易证

$$\frac{f_{\max}^n}{\sum\limits_{j \in S_0} (f(j))^n} = \frac{f_{\max}^n}{|S_1| f_{\max}^n + \sum\limits_{j \notin S_1} (f(j))^n} \leqslant \frac{1}{|S_1|} \tag{1.5.8}$$

于是得到

$$\sum_{i \in S} |P\{T_s(n) = i/A\} - \bar{a}_i|$$

$$= \sum_{i \notin S_1} |P\{T_s(n) = i/A\}| + \sum_{i \in S_1} \left| \frac{f_{\max}^n}{\sum\limits_{j \in S_0} (f(j))^n} - \frac{1}{|S_1|} \right|$$

$$= \sum_{i \notin S_1} |P\{T_s(n) = i/A\}|$$

$$+ |S_1| \cdot \left[ \frac{1}{|S_1|} - \frac{f_{\max}^n}{\sum\limits_{j \in S_0} (f(j))^n} \right]$$

$$= 1 + \sum_{i \notin S_1} |P\{T_s(n) = i/A\}| - \sum_{i \in S_1} |P\{T_s(n) = i/A\}|$$

$$= 2 - 2|S_1| \frac{f_{\max}^n}{\sum\limits_{j \in S_0} (f(j))^n}$$

**推论 1.5.1** 在定理 1.5.3 的假定下，有

$$\sum_{i \in S} |P\{T_s(n) = i/A\} - \bar{a}_i| \leqslant 2\left(\frac{N}{|S_1|} - 1\right)\lambda^{-n}$$

其中

$$\lambda = \min\left(\frac{f_{\max}}{f(i)}; i \in S_0\right)$$

**证明** 由定理 1.5.3 易得

$$\sum_{i \in S} |P\{T_s(n) = i/A\} - \bar{a}_i|$$

$$= 2 - 2|S_1| \frac{f_{\max}^n}{\sum_{j \in S_0}(f(j))^n} = 2 - \frac{2|S_1|}{\sum_{j \in S_0}\left(\frac{f(j)}{f_{\max}}\right)^n}$$

$$= 2 - \frac{2|S_1|}{|S_1| + \sum_{j \notin S_1}\left(\frac{f(j)}{f_{\max}}\right)^n}$$

$$\leqslant 2 - \frac{2|S_1|}{|S_1| + (|S_0| - |S_1|)\lambda^{-n}}$$

$$\leqslant \frac{2(|S_0| - |S_1|)\lambda^{-n}}{|S_1|} = 2\left(\frac{N}{|S_1|} - 1\right)\lambda^{-n}$$

推论1.5.1说明了收敛速度与$N$有关,$N$越小收敛速度越快。但是当$i \notin S_1$时有

$$\bar{a}_i = \lim_{n \to \infty} P\{T_s(n) = i/A\} = 0$$

于是$f$的最大值若落在$S_0$之外,通过适应值的选择算法则不可能得到全局最优值。

变异是一个重要的遗传机制。如果没有变异,遗传过程经常收敛到局部极值。在没有变异的情况下,遗传过程要达到最大值依赖于初始种群的性能。下面我们在不考虑选择的情况下研究变异的规律,即假定对于任意个体的$i \in S$有

$$P\{T_s(0) = i/A\} = a_i$$

**定理1.5.4**　设$A$是初始种群,对于任意$n \geqslant 1$有

$$P\{T_m(n) = i/A\}$$
$$= \sum_{j \in S} a_j \cdot \prod_{k=1}^{l} [0.5 + (\delta_{i_k j_k} - 0.5) \cdot (1 - 2\mu)^n] \tag{1.5.9}$$

其中$\mu$为变异概率。

**证明**　首先我们考虑变异使个体$i$转移到个体$j$的概率。由于$i$的分量的变异是独立的,且

$$P\{T_m(1)_k = j_k/i_k\} = \begin{cases} 1-\mu, & i_k = j_k \\ \mu, & i_k \neq j_k \end{cases}$$

于是

$$P\{T_m(1)_k = j_k/i_k\} = \mu + \delta_{i_kj_k}(1-2\mu) \qquad (1.5.10)$$

从而

$$P\{T_m(n+1) = j/T_m(n) = i\}$$
$$= \prod_{k=1}^{l}(\mu + \delta_{i_kj_k}(1-2\mu)) \qquad (1.5.11)$$

我们用归纳法来证明式(1.5.9)。当 $n = 1$ 时,由式(1.5.11) 即得

$$P\{T_m(1) = i_k/A\} = \sum_{j\in S}a_jP\{T_m(1) = i/j\}$$
$$= \sum_{j\in S}a_j \cdot \prod_{k=1}^{l}(\mu + \delta_{i_kj_k}(1-2\mu))$$
$$= \sum_{j\in S}a_j \cdot \prod_{k=1}^{l}(0.5 + (\delta_{i_kj_k} - 0.5)\cdot(1-2\mu))$$

于是式(1.5.9) 对于 $n = 1$ 成立。若式(1.5.9) 对于 $n$ 成立,则

$$P\{T_m(n+1) = z/A\}$$
$$= \sum_{i\in S}P\{T_m(n+1) = z/T_m(n) = i\}\cdot P\{T_m(n) = i/A\}$$
$$= \sum_{j\in S}a_j \cdot \prod_{k=1}^{l}\sum_{i_k}(\mu + \delta_{i_kz_k}(1-2\mu))$$
$$\cdot[0.5 + (\delta_{i_kj_k} - 0.5)\cdot(1-2\mu)^n]$$

我们将 $i_k, j_k, z_k$ 分以下情况:

(a) $j_k = z_k$,这时有 $j_k = i_k, j_k \neq i_k$ 两种情况;

(b) $j_k \neq z_k$,这时有 $z_k = i_k$ 和 $j_k = i_k, z_k \neq i_k$ 两种情况。

于是有

$$\sum_{i_k}(\mu + \delta_{i_kz_k}(1-2\mu))[0.5 + (\delta_{i_kj_k} - 0.5)\cdot(1-2\mu)^n]$$

$$
\begin{aligned}
&= \delta_{j_k z_k}[(1-\mu)(0.5+0.5(1-2\mu)^n) \\
&\quad + \mu\cdot(0.5-0.5(1-2\mu)^n)] \\
&\quad + (1-\delta_{j_k z_k})[(1-\mu)(0.5-0.5(1-2\mu)^n) \\
&\quad + \mu(0.5+0.5(1-2\mu)^n)] \\
&= \delta_{j_k z_k}(0.5+0.5(1-2\mu)^{n+1}) \\
&\quad + (1-\delta_{j_k z_k})(0.5-0.5(1-2\mu)^{n+1}) \\
&= 0.5+(\delta_{j_k z_k}-0.5)(1-2\mu)^{n+1}
\end{aligned}
$$

从而即得

$$
\begin{aligned}
&P\{T_m(n+1)=z/A\} \\
&= \sum_{j\in S} a_j \cdot \prod_{k=1}^{l}[0.5+(\delta_{j_k z_k}-0.5)\cdot(1-2\mu)^{n+1}]
\end{aligned}
$$

定理得证。

**推论 1.5.2**　设 $A$ 是初始无限种群，$\mu<0.5$，则

$$
\lim_{n\to\infty} P\{T_m(n)=i/A\} = \frac{1}{|S|} \tag{1.5.12}
$$

而且收敛是单调的，即

$$
\begin{aligned}
&\sum_{i\in S}\left|P\{T_m(n+1)=i/A\}-\frac{1}{|S|}\right| \\
&\leqslant \sum_{i\in S}\left|P\{T_m(n)=i/A\}-\frac{1}{|S|}\right|
\end{aligned} \tag{1.5.13}
$$

**证明**　由定理 1.5.4 及 $(1-2\mu)<1$，则证

$$
\begin{aligned}
&\lim_{n\to\infty} P\{T_m(n)=i/A\} = \sum_{j\in S} a_j \prod_{k=1}^{l}(0.5) \\
&= \sum_{j\in S} a_j\cdot(0.5)^l = \left(\frac{1}{2}\right)^l = \frac{1}{2^l} = \frac{1}{|S|}
\end{aligned}
$$

对于定理 1.5.4，若初始无限种群 $A'$ 为均匀分布，则对于任意 $n\geqslant 1$ 有

$$
P\{T_m(n)=i/A'\}
$$

$$= \frac{1}{|S|}\sum_{j\in S}\prod_{k=1}^{l}[0.5+(\delta_{i_kj_k}-0.5)(1-2\mu)^n]$$

$$= \frac{1}{|S|}\prod_{k=1}^{l}\sum_{j_k}[0.5+(\delta_{i_kj_k}-0.5)(1-2\mu)^n]$$

$$= \frac{1}{|S|}\prod_{k=1}^{l}[0.5+0.5(1-2\mu)^n+0.5-0.5(1-2\mu)^n]$$

$$= \frac{1}{|S|}$$

于是有

$$\sum_{i\in S}\left|P\{T_m(n+1)=i/A\}-\frac{1}{|S|}\right|$$

$$= \sum_{i\in S}\left|\sum_{j\in S}a_j(n)\cdot\prod_{k=1}^{l}[\mu+\delta_{i_kj_k}(1-2\mu)]-\frac{1}{|S|}\right|$$

$$= \sum_{i\in S}\left|\sum_{j\in S}\left(a_j(n)-\frac{1}{|S|}\right)\prod_{k=1}^{l}[\mu+\delta_{i_kj_k}(1-2\mu)]\right|$$

$$\leqslant \sum_{j\in S}\left|a_j(n)-\frac{1}{|S|}\right|\sum_{i\in S}\prod_{k=1}^{l}[(\mu+\delta_{i_kj_k}(1-2\mu))$$

$$= \sum_{j\in S}\left|a_j(n)-\frac{1}{|S|}\right|$$

$$= \sum_{j\in S}\left|P\{T_m(n)=j/A\}-\frac{1}{|S|}\right|$$

则证式(1.5.13)。

定理1.5.3及其推论说明了，不管初始无限种群给出什么分布，其极限分布都是均匀分布。也即是说，变异运算不断重复进行，可以达到个体空间中任何个体。变异把整个个体空间作为搜索空间。

我们进一步考虑杂交算子。取 $S$ 上的随机变量 $\hat{u}$，则

$$P\{T_c(n)=i/T_{s_1}(n)=j,T_{s_2}(n)=k\}$$

$$= P\{(j\otimes\hat{u})\oplus(k\otimes(1\oplus\hat{u}))\}$$

**例 1.5.2** 单点杂交 $\hat{u}$ 是下面集合上的均匀分布：

$$S_0 = \{(0\ 0 \cdots 0\ 1), (0\ 0 \cdots 0\ 1\ 1), \cdots, (1\ 1 \cdots 1)\}$$

$S_0$ 中只有 $l$ 个元素，即 $\hat{u}$ 在$S_0$ 上的均匀分布概率为$1/l$。单点杂交是在 $T_{s_1}(n)$ 中取右边部分，在 $T_{s_2}(n)$ 中取左边部分而构成新的后代。

**例 1.5.3** 两点杂交 $\hat{u}$ 是下面集合上的均匀分布：

$$\begin{aligned} S_2 = \{&(0 \cdots 0\ 1), (0 \cdots 0\ 1\ 0), \cdots, (1\ 0 \cdots 0), \\ &(0 \cdots 0\ 1\ 1), (0 \cdots 0\ 1\ 1\ 0), \cdots, (1\ 1\ 0 \cdots 0), \\ &\cdots \\ &\cdots, (1\ 1 \cdots 1)\} \end{aligned}$$

两点杂交是在中间集部分采用 $T_{s_1}(n)$，而在其它部分采用 $T_{s_2}(n)$。由于 $S_2$ 中有元素个数为

$$|S_2| = l + (l-1) + \cdots + 2 + 1 = \frac{l(l+1)}{2}$$

于是两点杂交 $\hat{u}$ 是$S_2$ 上分布概率为$\frac{2}{l(l+1)}$的均匀分布。

**例 1.5.4** 均匀随机杂交是按照分布

$$P\{\hat{u} = u\} = \mu^{|u|}(1-\mu)^{l-|u|}$$

当 $\mu = 0.5$ 时

$$P\{\hat{u} = u\} = \frac{1}{2^l}$$

表示均匀杂交。

**例 1.5.5** 半随机杂交，$\hat{u}$ 是$S_3$ 上的均匀分布。

$$S_3 = \begin{cases} \left\{u \in S, |u| = \frac{l}{2^l}\right\}, & l \text{ 为偶数} \\ \left\{u \in S, |u| = \frac{l-1}{2} \vee |u| = \frac{l+1}{2}\right\}, & l \text{ 为奇数} \end{cases}$$

这类杂交取每个双亲的一半基因杂交后产生后代。

**定理 1.5.5** 设 $\hat{u} = (\hat{u}_1, \cdots, \hat{u}_l)$ 是 $S = \{0,1\}^l$ 上的杂交随机变量，且对所有的 $k, m \in \{1,2,\cdots,l\}, k \neq m$，有

$$P\{\hat{u}_k \neq \hat{u}_m\} > 0$$

则

$$\lim_{n \to \infty} P\{T_c(n) = i/A\} = \prod_{k=1}^{l} P\{T_c(1)_k = i_k/A\} \tag{1.5.14}$$

其中 $A$ 为初始无限种群。收敛速度为

$$\sum_{i \in S} |P\{T_c(n) = i/A\} - \bar{a}_i| = C \cdot (\lambda(P_u, l))^n \tag{1.5.15}$$

其中

$$\lambda(P_u, l) = \max\{P\{\hat{u}_k \neq \hat{u}_m\}; k, m \leqslant l, k \neq m\} \tag{1.5.16}$$

**证明** 为了证明叙述简单起见，我们考虑单点杂交。不难证明

$$P\{T_c(1) = i/A\} = \sum_{(j \odot u) \oplus (m \odot \bar{u}) = i} a_j \cdot a_m \cdot P\{u\}$$

由于单点杂交是

$$S_1 = \{(0 \cdots 0\ 1), (0 \cdots 0\ 1\ 1), \cdots, (1\ 1 \cdots 1)\}$$

上的均匀分布，则

$$P\{T_c(1) = i/A\} = \frac{1}{l} \sum_{(j \odot u) \oplus (m \odot \bar{u}) = i} a_j \cdot a_m$$

$$= \frac{1}{l} \sum_{k=1}^{l} a_{i_1 \cdots i_k \cdots} a_{\cdots i_{k+1} \cdots i_l}$$

其中 $a_{i_1 \cdots i_k \cdots}$ 及 $a_{\cdots i_{k+1} \cdots i_l}$ 是 $a$ 在二进制表示下的边际分布。特别当 $A$ 是 $S$ 上的均匀分布时有

$$a_{i_1 \cdots i_k \cdots} = \frac{1}{2^k}$$

$$a_{\cdots i_{k+1} \cdots i_l} = \frac{1}{2^{l-k}}$$

于是

$$P\{T_c(1) = i/A\} = \frac{1}{2^l}$$

仍然为 $S$ 上的均匀分布，则可证

$$P\{T_c(n) = i\} = \frac{1}{2^l}$$

同样可证

$$P\{T_c(1)_k = i_k\} = \sum_{j_k = i_k} a_{j_1 \cdots j_l} = \frac{1}{2}$$

于是有

$$P\{T_c(n) = i/A\} = \prod_{k=1}^{l} P\{T_c(1)_k = i_k/A\}$$

式(1.5.14)显然成立。

对于单点杂交，$k < m$，计算

$$P\{\hat{u}_k \neq \hat{u}_m\} = P\{\hat{u}_k = 0, \hat{u}_m = 1\}$$

显然有

$$\begin{aligned} P\{\hat{u}_1 = 0, \hat{u}_l = 1\} &= P\{\hat{u} = (0\,0 \cdots 1)\} + P\{\hat{u} = (0 \cdots 1\,1)\} \\ &\quad + \cdots + P\{\hat{u} = (0\,1 \cdots 1)\} \\ &= \frac{l-1}{l} = 1 - \frac{1}{l} \end{aligned}$$

对于其它情况，均有 $P\{\hat{u}_k \neq \hat{u}_m\} < 1 - 1/l$，于是

$$\lambda(P_u, l) = 1 - \frac{1}{l} \tag{1.5.17}$$

对于两点杂交有

$$\lambda(P_u, l) = 1 - \frac{2}{l-1} \tag{1.5.18}$$

对于随机均匀杂交有

$$\lambda(P_u, l) = 1 - 2\mu + 2\mu^2 \tag{1.5.19}$$

特别对于均匀杂交有

$$\lambda(P_u, l) = 0.5 \tag{1.5.20}$$

如果记

$$\lambda_l = \min_{P_u} \lambda(P_u, l) \tag{1.5.21}$$

其中 $p_u$ 是所有可能的杂交，则对于半随机杂交有

$$\lambda(P_u, l) = \lambda_l \tag{1.5.22}$$

**定理1.5.6** 设 $l$ 是偶数，$S = \{0,1\}^l$，$\hat{u}$ 为杂交算子，对于以下三个命题：

(1) $\lambda(P_u, l) = \lambda_l$；

(2) 对于任意 $k, m \leqslant l, k \neq m, P\{\hat{u}_k = \hat{u}_m\} = \lambda_l$；

(3) $P\left\{|\hat{u}| \neq \frac{l}{2}\right\} = 0$。

有关系：

(1)⇔(2)⇒(3)。

**证明** 首先计算

$$\sum_{k \neq l} P\{\hat{u}_k = \hat{u}_l / \hat{u} = u\} = \frac{|u| \cdot (|u| - 1)}{2} + \frac{(l - |u|)(l - |u| - 1)}{2}$$

$$= \frac{l \cdot (l-1)}{2} - |u| \cdot (l - |u|) \geqslant \frac{l(l-2)}{4} \tag{1.5.23}$$

于是有

$$\frac{l(l-1)}{2} \cdot \lambda(P_u, l) \geqslant \sum_{k \neq m} P\{\hat{u}_k = \hat{u}_m\}$$

$$= \sum_{u\in S}\sum_{k\neq m} P\{\hat{u}_k = \hat{u}_m / \hat{u} = u\} \cdot P\{\hat{u} = u\}$$
$$\geqslant \min\left\{\sum_{k\neq m} P\{\hat{u}_k = \hat{u}_m / \hat{u} = u\}; u \in S\right\}$$
$$= \frac{l(l-2)}{4} \tag{1.5.24}$$

(1)⇒(2)：若(1) 成立，即 $\lambda(P_u, l) = \lambda_l = \min\limits_{u}\lambda(P_{\hat{u}}, l)$ 式(1.5.24) 的不等式变为等式。而 $\sum\limits_{k\neq m}$ 只有 $\frac{l(l-1)}{2}$ 项，于是(2) 必须成立。

(2)⇒(1)：由 $\lambda_l$ 的定义可证。

(1)⇒(3)：若(1) 成立，则 $\sum\limits_{k\neq m} P\{\hat{u}_k = \hat{u}_m / \hat{u} = u\} = c$，对于所有 $u \in S$ 成立，则证(3) 成立。

定理 1.5.5 和定理 1.5.6 说明了杂交算子的性质。杂交算子反复运算的结果依赖于初始分布。

选择、变异和杂交机制各有其功能。单纯利用选择可找到局部最优值，变异可以使搜索遍及整个空间，而杂交结果依赖于初始分布。因此，一个完整的遗传算法应当是选择、变异和杂交运算共同构成的。

## 1-6　遗传算法的几何解释

假定 $T_s$ 为选择算子，$T_c$ 为杂交算子，$T_m$ 为变异算子，记

$$G(p) = T_m \circ T_c \circ T_s(p) \tag{1.6.1}$$

其中 $p$ 为 $S$ 上的无限种群，即

$$p \in \Lambda = \left\{(a_0, \cdots, a_{M-1});\ a_i \geqslant 0,\ \sum_{i=0}^{M-1} a_i = 1\right\}$$

则 $G$ 为 $\Lambda$ 到 $\Lambda$ 的映射。

遗传算法是固定长度 $N$ 的有限种群。例如取 $S=\{0,1\}^3=\{0,1,\cdots,7\}$，且初始种群

$$X_0=(0\ 1\ 1\ 0\ 2\ 1\ 2\ 3\ 4\ 3)$$

则初始种群分布

$$p_0=(0.2\ 0.3\ 0.2\ 0.2\ 0.1\ 0\ 0\ 0)$$

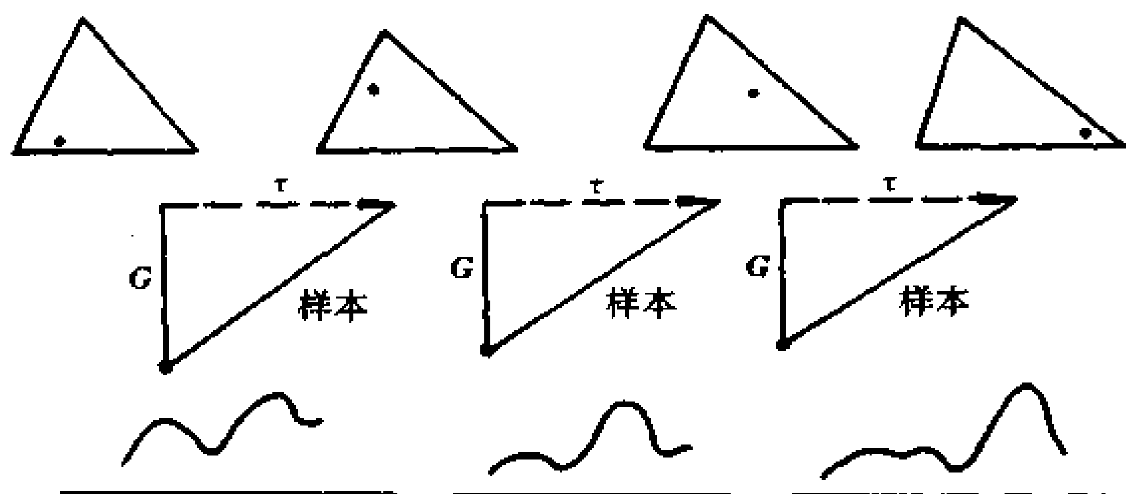



图 1.6.1 遗传算法演化过程

由于 $X_0$ 为有限种群，按照 $\boldsymbol{G}(p)$ 提取长度为 10 的样本即得到下一代种群，比如

$$X_1=(1\ 0\ 2\ 0\ 1\ 3\ 2\ 4\ 3\ 4)$$

于是得到

$$p_1=(0.2\ 0.2\ 0.2\ 0.2\ 0.2\ 0\ 0\ 0)$$

如果我们把 $\Lambda$ 看作 $S$ 中的 $(n-1)$ 维凸多面体，则遗传算法过程可视为图 1.6.1 的过程。

如果我们假定种群规模为 $N$，则种群变化范围为 $\Lambda$ 中的有理点，即在下面的范围中变化：

$$\frac{1}{N}\Lambda_M^N=\frac{1}{N}\left\{(N_0\cdots N_{M-1});\ \sum_{i=0}^{M-1}N_i=N\right\}$$

**定理 1.6.1** 设 $p$ 为当前种群，则对长度 $N$ 的有限种群，其下

代种群 $\tau(p)$ 的期望值为 $G(p)$。

**证明**　为了得到下一代种群

$$q = \left(\frac{N_0}{N}, \cdots, \frac{N_{M-1}}{N}\right)$$

则按照 $G(p)$ 抽样，即得

$$p_0' = \binom{N}{N_0}(G(p)_0)^{N_0}$$

$$p_1' = \binom{N - N_0}{N_1}(G(p)_1)^{N_1}$$

$$\vdots$$

$$p_{M-1}' = \binom{N - N_0 - \cdots - N_{M-2}}{N_{M-1}}(G(p)_{M-1})^{N_{M-1}}$$

$q$ 发生的概率为 $p_0' \cdots p_{M-1}'$，即

$$N! \prod_{j<M} \frac{(G(p)_j)^{N_j}}{N_j!}$$

于是 $p_i'$ 的期望值为

$$N! \sum_{q \in \Lambda_M^N / N} \frac{N_i}{N} \prod_{j<M} \frac{(G(p)_j)^{N_j}}{N_j!}$$

根据多项式定理有

$$\left(\sum_{i=1}^{M-1} x_i\right)^N = N! \sum_{N_0 + \cdots + N_{M-1} = N} \prod_{j<M} \frac{x_j^{N_j}}{N_j!}$$

两边对 $x_i$ 求导即得

$$N x_i \left(\sum_{i=0}^{M-1} x_i\right)^{N-1} = N! \sum_{N_0 + \cdots + N_{n-1} = N} N_i \prod_{j<M} \frac{x_j^{N_j}}{N_j!}$$

于是 $p_i'$ 发生的期望值为

$$(G(p))_i \left(\sum_{i=0}^{M-1} G(p)_i\right)^{N-1} = (G(p))_i$$

则证。

由此可见，在无限种群的遗传机制中，若给出初始种群 $p$，则得到遗传序列

$$p, \quad G(p), \quad G^2(p), \quad \cdots$$

当种群为有限值时，$G(p)$ 可以得到样本分布；由于样本是多样的，其期望值为 $G(p)$。由于定理 1.6.1 与样本大小 $N$ 无关，所以 $G$ 描述了所有种群规模的变化规律。

**定义 1.6.1**　若对于任意 $j$ 和 $k$ 有

$$\mu_j = \sum_{k\otimes i=0} \mu_{i\oplus j} \sum_{\bar{k}\otimes i=0} \mu_{i\oplus j} \tag{1.6.2}$$

则称变异为独立的。若

$$\mu_i = \mu^{T_i}(1-\mu)^{l-T_i} \tag{1.6.3}$$

$T_i$ 表示 $i$ 中 1 的个数，$\mu_i$ 表示将任意 $j \in S$ 变为 $j \oplus i$ 的概率。$\mu$ 被称为变异率。

**定理 1.6.2**　变异率 $\mu$ 是独立的。

**证明**　首先有

$$\sum_{k\otimes i=0} \mu_{i\oplus j} = \sum_{\bar{k}\otimes(i\oplus j)=0} \mu_i = \sum_{\bar{k}\otimes i=\bar{k}\otimes j} \mu^{T_i}(1-\mu)^{l-T_i}$$

$$= (1-\mu)^l \sum_{\bar{k}\otimes i=\bar{k}\otimes j} \left(\frac{\mu}{1-\mu}\right)^{T_i}$$

由于 $i = \bar{k}\otimes j \oplus k \otimes u$ 时 $\bar{k}\otimes i = \bar{k}\otimes j$，于是上述在 $i$ 上求和相当于在 $u \in S_k = \{i;\ i\otimes\bar{k} = 0\}$ 上求和，即

$$\sum_{\bar{k}\otimes i=0} \mu_{i\oplus j} = (1-\mu)^l \sum_{u\in S_k} \left(\frac{\mu}{1-\mu}\right)^{T_{\bar{k}\otimes j}+T_u}$$

$$= (1-\mu)^{l-T_k} \left(\frac{\mu}{1-\mu}\right)^{T_{\bar{k}\otimes j}} \sum_{u\in S_k} \mu^{T_u}(1-\mu)^{T_k-T_u}$$

$$= (1-\mu)^{l-T_k} \left(\frac{\mu}{1-\mu}\right)^{T_{\bar{k}\otimes j}}$$

于是易证

$$\sum_{k\otimes i=0} \mu_{i\oplus j} \sum_{\bar{k}\otimes i=0} \mu_{i\oplus j}$$

$$= (1-\mu)^{l-T_k}\left(\frac{\mu}{1-\mu}\right)^{T_{k\otimes j}}(1-\mu)^{l-T_{\bar{k}}}\left(\frac{\mu}{1-\mu}\right)^{T_{\bar{k}\otimes j}}$$

$$= \mu_j$$

**定理 1.6.3**　如果变异是独立的,则

$$P\{T_c(T_m(x),T_m(y)) = z\} = P\{T_m(T_c(x,y)) = z\} \tag{1.6.4}$$

对于任意 $x,y,z \in S$ 成立。也即是说,先变异后杂交与先杂交后变异有相同的分布。

**证明**　计算

$$P\{T_c(T_m(x),T_m(y)) = z\} = \sum_{i,j,k\in M}\mu_i\mu_j \cdot \frac{p_k + p_{\bar{k}}}{2}$$

其中

$$\begin{aligned} M &= \{(i,j,k);\ (x\oplus i)\otimes k\oplus(y\oplus j)\otimes\bar{k} = z\} \\ &= \{(i,j,k);\ (x\otimes k)\oplus y\otimes\bar{k} = z\oplus b, \\ &\qquad b = i\otimes k\oplus j\otimes\bar{k}\} \end{aligned}$$

于是

$$\begin{aligned} &P\{T_c(T_m(x),T_m(y)) = z\} \\ &\quad = \sum_k \frac{p_k + p_{\bar{k}}}{2}\sum_{x\otimes k\oplus y\otimes\bar{k} = z\oplus b}\ \sum_{i\otimes k\oplus j\otimes\bar{k} = b}\mu_i\mu_j \end{aligned}$$

又因为

$$\begin{aligned} &i\otimes k\oplus j\otimes\bar{k} = b = k\otimes b\oplus\bar{k}\otimes b \\ &\Leftrightarrow k\otimes i = k\otimes b \text{ 且 } \bar{k}\otimes j = \bar{k}\otimes b \end{aligned}$$

由 $\mu$ 的独立性有

$$\begin{aligned} \sum_{i\otimes k\oplus j\otimes\bar{k} = b}\mu_i\mu_j &= \sum_{k\otimes(i\oplus b)=0}\mu_i\sum_{\bar{k}\otimes(j\oplus b)=0}\mu_j \\ &= \sum_{k\otimes i=0}\mu_{b\oplus i}\sum_{\bar{k}\otimes j=0}\mu_{b\oplus j} = \mu_b \end{aligned}$$

于是

$$P\{T_c(T_m(x),T_m(y)) = z\}$$

$$= \sum_k \frac{p_k + p_{\bar{k}}}{2} \sum_{x \otimes k \oplus y \otimes \bar{k} = z \oplus b} \mu_b$$

$$= \sum_b \mu_b \sum_{x \otimes k \oplus y \otimes \bar{k} = z \oplus b} \frac{p_k + p_{\bar{k}}}{2}$$

$$= P\{T_m(T_c(x, y)) = z\}$$

**定理 1.6.4** 用 $m_{x,y}(z)$ 表示 $x, y$ 通过杂交与变异而得到 $z$ 的概率,则先变异后杂交的概率为

$$m_{x,y}(z) = \sum_{i,j,k} \mu_i \mu_j \frac{p_k + p_{\bar{k}}}{2} \cdot \chi_{\{(x \oplus i) \otimes k \oplus (y \oplus j) \otimes \bar{k} = z\}} \tag{1.6.5}$$

而先杂交后变异的概率为

$$m_{xy}(z) = \sum_{j,k} \mu_j \frac{p_k + p_{\bar{k}}}{2} \cdot \chi_{\{x \otimes k \oplus y \otimes \bar{k} = z \oplus j\}} \tag{1.6.6}$$

**证明** 由定义直接得到。

**定理 1.6.5** 无论变异是否独立,杂交与变异孰先孰后,均有

$$m_{xy}(z) = m_{yx}(z) = m_{x \oplus z, y \oplus z}(0) \tag{1.6.7}$$

**证明** 由定理 1.6.4 知,$x$ 与 $y$ 是对称的。又因为

$$(x \oplus i) \otimes k \oplus (y \oplus j) \otimes \bar{k} = z$$

等价于

$$(x \oplus i \oplus z) \otimes k \oplus (y \oplus j \oplus z) \otimes \bar{k} = 0$$

则证先变异后杂交时有

$$m_{x,y}(z) = m_{y,x}(z) = m_{x \oplus z, y \oplus z}(0)$$

对于先杂交后变异有类似证明。

记

$$\boldsymbol{M} = (m_{ij}(0);\ i, j \in S)$$

则 $\boldsymbol{M}$ 为对称矩阵。记

$$(\sigma_k)_{ij} = \chi_{\{i \oplus j = k\}} = \delta_{i \oplus j, k}$$

$$\sigma_k = ((\sigma_k)_{ij};\ i, j \in S)$$

于是

$$\sigma_k \cdot (a_0, \cdots, a_{M-1})^{\mathrm{T}} = (a_{0 \oplus k}, \cdots, a_{(M-1) \oplus k})^{\mathrm{T}}$$

容易验证 $\sigma_k$ 是对称的,且

$$\sigma_i \cdot \sigma_j = \sigma_{i \oplus j} = \sigma_j \cdot \sigma_i$$

取映射 $\mathscr{M}: \Lambda \to \Lambda$ 为

$$\mathscr{M}(a)_i = a^{\mathrm{T}} \sigma_i M \sigma_i a$$

以及映射 $\mathscr{F}: \Lambda \to \Lambda$ 为

$$\mathscr{F}_i(p) = P\{T_s(p) = i\}$$

于是得到复合映射 $\mathscr{M} \circ \mathscr{F}: \Lambda \to \Lambda$

**定理 1.6.6**　对于标准遗传算法有

$$G(p) = \mathscr{M} \circ \mathscr{F}(p) \ (p \in \Lambda) \tag{1.6.8}$$

**证明**　由于

$$\begin{aligned}
G_i(p) &= P\{T_c(T_m(T_s(p)), T_m(T_s(p))) = i\} \\
&= \sum_{x,y} P\{(T_s(p), T_s(p)) = (x, y)\} \\
&\quad \cdot P\{i = T_c(T_m(x), T_m(y))\} \\
&= \sum_{x,y} \mathscr{F}(p)_x \mathscr{F}(p)_y \sum_{u,v} \mu_{x \oplus u} \mu_{y \oplus v} P\{i = T_c(u, v)\} \\
&= \sum_{x,y} \mathscr{F}(p)_x \mathscr{F}(p)_y \sum_{u,v} \mu_{x \oplus u} \mu_{y \oplus v} \sum_{u \otimes k \oplus v \otimes \bar{k} = i} \frac{p_k + p_{\bar{k}}}{2} \\
&= \sum_{x,y} \mathscr{F}(p)_x \mathscr{F}(p)_y \sum_{u,v,k} \mu_u \mu_v \frac{p_k + p_{\bar{k}}}{2} \chi_{\{(x \oplus u) \otimes k \oplus (y \oplus v) \otimes \bar{k} = i\}} \\
&= \sum_{x,y} \mathscr{F}(p)_x \mathscr{F}(p)_y m_{x \oplus i, y \oplus i}(0) = (\mathscr{M} \circ \mathscr{F}(p))_i
\end{aligned}$$

则证。

**定义 1.6.2**　如果对于任意 $p \in \Lambda$ 有

$$\lim_{n \to \infty} G^n(p) = p' \in \Lambda$$

称 $G$ 是收敛的。这时

$$G(p') = G(\lim_{n \to \infty} G^n(p)) = \lim_{n \to \infty} G^{n+1}(p) = p'$$

称 $p'$ 为 $G$ 的不动点。

**定理 1.6.7** 设 $G$ 有有限个不动点，$\varphi$ 是连续函数，且满足

$$x \neq G(x) \Rightarrow \varphi(x) > \varphi(G(x)) \tag{1.6.9}$$

则 $G$ 是收敛的。

**证明** 设 $\omega_j (j \leqslant N_0)$ 为 $\Lambda$ 中关于 $G$ 的有限不动点，记

$$U_j = \{x \in \Lambda;\ \| x - \omega_j \| < \epsilon\} \qquad (j \in N_0)$$

取 $\epsilon$ 充分小，使 $x \in U_j$ 时 $G \notin U_k (k \neq j, k \leqslant N_0)$。利用反证法来证明。假定 $G$ 不是收敛的，由上述假定知 $\{G^n(x)\}$ 必有无限次落入 $U = \bigcup_{j=1}^{N_0} U_j$ 的补集 $\overline{U}$。于是存在 $\{G^n(x)\}$ 的极限点 $z \notin U$。又因为 $\varphi(z) > \varphi(G(z))$，由 $\varphi$ 的连续性，存在 $\epsilon' > 0$ 使对于 $V = \{y \in \Lambda;\ \| y - z \| \leqslant \epsilon\}$ 有

$$\varphi(z) > \varphi(G(V)) = \max_{y \in V} \varphi(G(y))$$

从而 $\varphi(z) > \varphi(G(V)) \geqslant \varphi(G^2(V)) \cdots$ 但是 $z$ 作为 $\{G^n(x)\}$ 的极限点，必落在 $\bigcup_k G^k(V)$ 中，于是 $\varphi(z) > \varphi(z)$，矛盾。

我们指出，条件式(1.6.9)类似于 Lyapunov 函数。式(1.6.9)也可改为

$$x \neq G(x) \Rightarrow \varphi(x) < \varphi(G(x))$$

**定义 1.6.3** 我们称适应值函数为线性的，若存在 $a \in R^l$，$b \in R$，使得

$$f(i) = a^{\mathrm{T}} i + b = \sum_{k=0}^{l-1} a_k i_k + b$$

线性适应值函数相当于对基因分量加权。特别当 $b = 0, a_k = 2^k$ 时，$f(i)$ 即是将二进制变为十进制。

为了讨论线性适应值函数的不动点，先证明以下性质。

**定理 1.6.8** 对于遗传机制有以下性质：

$$(1)\ a^{\mathrm{T}} i = \sum_{j=0}^{l-1} a_j i_j = \frac{1}{2} \sum_{j=0}^{l-1} a_j (1 - (-1)^{i^{\mathrm{T}} \cdot 2^j}) \tag{1.6.10}$$

(2) $\sum_b (2m_{u\oplus b,v\oplus b}(0) - \delta_{ub} - \delta_{vb}) = 0$ (1.6.11)

(3) 若 $\mu = 0$,则

$$\sum_i (-1)^{i^T \cdot 2^j} m_{u\oplus i,v\oplus i}(0) = (-1)^{u^T \cdot 2^j} \chi_{(u\cdot 2^j = v\cdot 2^j)}$$

**证明** (1) 由于 $i^T \cdot 2^j = i_j$,当 $i_j \in \{0,1\}$ 时,有$(1-(-1)^{i_j}) = 2i_j$,则证(1)。

(2) 由于

$$\sum_b (2m_{u\oplus b,v\oplus b}(0) - \delta_{ub} - \delta_{vb})$$

$$= 2\sum_b \sum_{i,j,k} \mu_i\mu_j \frac{p_k + p_{\bar{k}}}{2} \chi_{\{(u\oplus b\oplus i)\otimes k\oplus(v\oplus b\oplus j)\otimes \bar{k} = 0\}} - \delta_{u,b} - \delta_{v,b}$$

$$= -2 + 2\sum_{i,j,k} \mu_i\mu_j \frac{p_k + p_{\bar{k}}}{2} \cdot \sum_b \chi_{\{b\otimes k\oplus b\otimes \bar{k}\oplus(u\oplus i)\otimes k\oplus(v\oplus j)\otimes \bar{k} = 0\}}$$

$$= -2 + 2\sum_{i,j,k} \mu_i\mu_j \frac{p_k + p_{\bar{k}}}{2} \sum_b \chi_{\{(u\oplus i)\otimes k\oplus(v\oplus j)\otimes \bar{k} = b\}}$$

$$= -2 + 2\sum_{i,j,k} \mu_i\mu_j \frac{p_k + p_{\bar{k}}}{2} = -2 + 2 = 0$$

(3) 利用 $i \oplus u$ 代替 $i$ 及 $u$ 与 $v$ 的对称性易得

$$\sum_i (-1)^{i^T \cdot 2^j} m_{u\oplus i,v\oplus i}(0)$$

$$= (-1)^{u^T \cdot 2^j} \sum_i (-1)^{i^T \cdot 2^j} m_{i,u\oplus v\oplus i}(0)$$

$$= (-1)^{v^T \cdot 2^j} \sum_i (-1)^{i^T \cdot 2^j} m_{i,u\oplus v\oplus i}(0)$$

$$= (-1)^{u^T \cdot 2^j} \chi_{\{u\otimes 2^j = v\otimes 2^j\}} \cdot \sum_i (-1)^{i^T \cdot 2^j} m_{i,u\oplus v\oplus i}(0)$$

$$= (-1)^{u^T \cdot 2^j} \chi_{\{u\otimes 2^j = v\otimes 2^j\}}$$

$$\cdot \sum_i (-1)^{i^T \cdot 2^j} \sum_k \frac{p_k + p_{\bar{k}}}{2} \chi_{\{i \odot k \oplus (u \oplus v \oplus i) \odot \bar{k} = 0\}}$$

$$= (-1)^{u^T \cdot 2^j} \chi_{\{u \odot 2^j = v \odot 2^j\}}$$

$$\cdot \sum_k \frac{p_k + p_{\bar{k}}}{2} \sum_i (-1)^{i^T \cdot 2^j} \chi_{\{(u \oplus v) \odot \bar{k} = i\}}$$

$$= (-1)^{u^T \cdot 2^j} \sum_k \frac{p_k + p_{\bar{k}}}{2} \chi_{\{u \odot 2^j = v \odot 2^j\}}$$

$$\cdot (-1)^{((u \oplus v) \odot \bar{k})^T \cdot 2^j}$$

$$= (-1)^{u^T \cdot 2^j} \chi_{\{u \odot 2^j = v \odot 2^j\}} \cdot \sum_k \frac{p_k + p_{\bar{k}}}{2}$$

**定理1.6.9** 如果适应值是线性的，变异率 $\mu = 0$，则 $G$ 是收敛的，不动点必在 $\Lambda$ 的顶点上，种群平均适应值是单调增的。

**证明** 令 $q = G(p)$，则

$$(f^T p)^2 q_i = (f^T p)(f_i p_i) + \sum_{u,v} f_u p_u f_v p_v (m_{u \oplus i, v \oplus i}(0) - \delta_{v,i})$$

于是

$$(f^T p)^2 f^T q = (f^T p) \sum_i f_i^2 p_i + \sum_{u,v} f_u p_u f_v p_v \sum_i f_i (m_{u \oplus i, v \oplus i}(0) - \delta_{v,i})$$

利用不等式

$$\sum_i f_i^2 p_i \geqslant \left( \sum_i f_i p_i \right)^2 \tag{1.6.12}$$

即得

$$f^T q \geqslant f^T p + s^T A s \tag{1.6.13}$$

其中

$$s = \mathcal{F}(p) = \frac{f \cdot p}{f^T p}$$

和

$$A_{uv} = \sum_i f_i (m_{u\oplus i, v\oplus i}(0) - \delta_{v,i})$$

利用定理1.6.8即证

$$A_{uv} = \frac{1}{2}\sum_i (b + a^{\mathrm{T}} i)(2m_{u\oplus i, v\oplus i}(0) - \delta_{v,i})$$

$$= \sum_i \frac{1}{4}\sum_j a_j (1 - (-1)^{i^{\mathrm{T}}\cdot 2^j})$$

$$\cdot (2m_{u\oplus i, v\oplus i}(0) - \delta_{u,i} - \delta_{v,i})$$

$$= \frac{-1}{4}\sum_i \sum_j a_j (-1)^{i^{\mathrm{T}}\cdot 2^j}(2m_{u\oplus i, v\oplus i}(0) - \delta_{u,i} - \delta_{v,i})$$

$$= \frac{-1}{4}\sum_j a_j \sum_i (-1)^{i^{\mathrm{T}}\cdot 2^j}(2m_{u\oplus i, v\oplus i}(0) - \delta_{u,i} - \delta_{v,i})$$

$$= \frac{-1}{4}\sum_j a_j [2(-1)^{u^{\mathrm{T}}\cdot 2^j}\chi_{(u\otimes 2^j = v\otimes 2^j)}$$

$$- (-1)^{u^{\mathrm{T}}\cdot 2^j} - (-1)^{v^{\mathrm{T}}\cdot 2^j}] = 0$$

由式(1.6.13)即得

$$\varphi(\boldsymbol{G}(p)) = f^{\mathrm{T}}(\boldsymbol{G}(p)) = f^{\mathrm{T}} q \geqslant f^{\mathrm{T}} p = \varphi(p) \tag{1.6.14}$$

(1.6.14)等式成立当且仅当式(1.6.12)中等式成立，即

$$\sum_i f_i^2 p_i = \left(\sum_i f_i p_i\right)^2 \tag{1.6.15}$$

由于

$$D(f) = \sum_i f_i^2 p_i - \left(\sum_i f_i p_i\right)^2$$

表示取值 $f_i$ 的概率为 $p_i$ 的方差，于是 $D(f) = 0$ 当且仅当 $p$ 为单点分布，即 $p$ 在 $\Lambda$ 的顶点上。当 $p$ 不在 $\Lambda$ 的顶点上时，$f^{\mathrm{T}} q > f^{\mathrm{T}} p$，于是 $p \neq q$ 时有 $\varphi(p) < \varphi(q)$，于是 Lyapunov 函数成立，故 $\boldsymbol{G}$ 必有不动点。由式(1.6.14)知种群平均适应值是在不断地增加的。

定理1.6.9指出对于线性适应值以及0变异率，$\boldsymbol{G}$ 是收敛的，且不动点存在。$\boldsymbol{G}$ 的收敛点必在 $\Lambda$ 的顶点上。

# 第 2 章

## 遗传算法的马氏链模型

遗传算法是不断重复杂交、变异和选择的过程。每一种遗传机制都与当前种群状态有关,而与以前的种群状态无关。因此遗传算法是一个马尔可夫链。

### 2-1 马尔可夫链的定义及性质

**定义 2.1.1** 设$\{X_n;n\geqslant 0\}$为一列取值离散的随机变量,离散值的全体记为$S=\{j\}$,称$S$为状态空间。若对于任意$n\geqslant 1$,$i_k\in S(k\leqslant n+1)$有

$$\begin{aligned} &P\{X_{n+1}=i_{n+1}/X_n=i_n,\cdots,X_0=i_0\}\\ &\quad =P\{X_{n+1}=i_{n+1}/X_n=i_n\} \end{aligned} \tag{2.1.1}$$

称$\{X_n;n\geqslant 0\}$为马尔可夫链。

(2.1.1) 式成立是指，只要等式两边的条件概率有意义，等式即成立。(2.1.1) 式即是一种无后效性，它大大简化了条件概率的计算。对于马尔可夫链有：

$$
\begin{aligned}
&P\{X_{n+2} = i_{n+2}, X_{n+1} = i_{n+1}/X_n = i_n\} \\
&\quad = P\{X_{n+1} = i_{n+1}/X_n = i_n\} \\
&\qquad \cdot P\{X_{n+2} = i_{n+2}/X_{n+1} = i_{n+1}, X_n = i_n\} \\
&\quad = P\{X_{n+1} = i_{n+1}/X_n = i_n\} \cdot P\{X_{n+2} = i_{n+2}/X_{n+1} = i_{n+1}\}
\end{aligned}
$$

**例 2.1.1**　设 $\{\xi_n; n \geqslant 0\}$ 为一列取整数值的随机变量，记部分和为

$$X_n = \sum_{k=0}^{n} \xi_k \qquad (n \geqslant 0)$$

若 $\{\xi_n; n \geqslant 0\}$ 是独立同分布的，即

$$P\{\xi_n = k\} = a_k$$

则 $\{X_n; n \geqslant 0\}$ 为马尔可夫链。这是因为在 $X_n = i_n$ 的条件下，$X_{n+1} = i + \xi_{n+1}$ 只与 $\xi_{n+1}$ 有关，而与 $X_{n-1}, X_{n-2}, \cdots$ 无关，因此

$$
\begin{aligned}
&P\{X_{n+1} = i_{n+1}/X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0\} \\
&\quad = P\{\xi_{n+1} = i_{n+1} - i_n\} = a_{i_{n+1} - i_n}
\end{aligned}
$$

**例 2.1.2**　设某种生物细胞中有 $N$ 对染色体，其中有一部分 $a$ 型的，另一部分是 $b$ 型的。在它繁殖过程中，如果父代的 $2N$ 个染色体中有 $j$ 个 $a$ 型，$(2N - j)$ 个 $b$ 型，那么子代每对染色体是 $a$ 型的概率是 $j/2N$，是 $b$ 型的概率为 $(2N - j)/2N$。记 $n$ 代的细胞中含 $a$ 型的个数为 $X_n$，则 $\{X_n; n \geqslant 0\}$ 是一个马尔可夫链。这时

$$P\{X_{n+1} = j/X_n = i\} = \binom{2N}{j}\left(\frac{i}{2N}\right)^j\left(\frac{2N - i}{2N}\right)^{2N-j}$$

在马尔可夫链的讨论中，我们看到条件概率

$$_m p_{ij}^{(n)} = P\{X_{n+m} = j/X_m = i\}$$

起着特别重要的作用。它表示系统在 $m$ 时刻处于状态 $i$，经过 $n$ 步转移达到状态 $j$ 的概率。一般来说，这个转移概率与初始时刻 $m$ 无关。即

$$p_{ij}^{(n)} = P\{\hat{X}_{n+m} = j/\hat{X}_m = i\}$$

这时称马尔可夫链为齐次的。

**定理 2.1.1** 齐次马尔可夫链的联合分布由初始分布

$$P\{\hat{X}_0 = i\} = p_i \qquad (i \in S)$$

及一步转移概率

$$p_{ij} = P\{\hat{X}_n = j/\hat{X}_{n-1} = i\} \qquad (i,j \in S)$$

完全确定。

**证明** 首先由马尔可夫链的定义有

$$\begin{aligned} &P\{\hat{X}_1 = i_1,\cdots,\hat{X}_n = i_n/\hat{X}_0 = i_0\} \\ &\quad = P\{\hat{X}_1 = i_1/\hat{X}_0 = i_0\}P\{\hat{X}_2 = i_2/\hat{X}_1 = i_1,\hat{X}_0 = i_0\} \\ &\qquad \cdots P\{\hat{X}_n = i_n/\hat{X}_0 = i_0,\cdots,\hat{X}_{n-1} = i_{n-1}\} \\ &\quad = P\{\hat{X}_1 = i_1/\hat{X}_0 = i_0\}P\{\hat{X}_2 = i_2/\hat{X}_1 = i_1\} \\ &\qquad \cdots P\{\hat{X}_n = i_n/\hat{X}_{n-1} = i_{n-1}\} \end{aligned}$$

则证

$$\begin{aligned} &P\{\hat{X}_0 = i_0,\cdots,\hat{X}_1 = i_1,\cdots,\hat{X}_n = i_n\} \\ &\quad = P\{\hat{X}_0 = i_0\}P\{\hat{X}_1 = i_1,\cdots,\hat{X}_n = i_n/\hat{X}_0 = i_0\} \\ &\quad = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \end{aligned}$$

**定理 2.1.2** 对于齐次马尔可夫链及任意 $n, m \geqslant 0, i, j \in S$ 有

$$p_{ij}^{(n+m)} = \sum_{k\in S} p_{ik}^{(n)} p_{kj}^{(m)} \tag{2.1.2}$$

**证明** 由概率的性质有

$$\begin{aligned} p_{ij}^{(n+m)} &= P\{\hat{X}_{n+m} = j/\hat{X}_0 = i\} \\ &= \sum_{k\in S} P\{\hat{X}_{n+m} = j/\hat{X}_n = k,\hat{X}_0 = i\}P\{\hat{X}_n = k/\hat{X}_0 = i\} \end{aligned}$$

$$= \sum_{k \in S} P\{X_n = k / X_0 = i\} P\{X_{n+m} = j / X_n = k\}$$

$$= \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}$$

式(2.1.2)称为C-K方程,即切普曼-柯尔莫哥洛夫方程。

假定状态集 $S$ 中状态是有限的,称马尔可夫链为有限的。记一次转移概率为矩阵形式,即

$$\boldsymbol{P} = (p_{ij};\ i,j = 0(n-1))$$

则C-K方程可以表示为

$$\boldsymbol{P}^2 = (p_{ij}^{(2)};\ i,j = 0(n-1)) = \boldsymbol{P} \cdot \boldsymbol{P}$$

一般地有

$$\boldsymbol{P}^{n+m} = \boldsymbol{P}^n \cdot \boldsymbol{P}^m$$

若初始分布为

$$p_i^{(0)} = P\{X_0 = i\}$$

记

$$p^{(0)} = (p_0^{(0)}, \cdots, p_{n-1}^{(0)})$$

则 $X_n$ 的分布为 $p^{(0)} \cdot \boldsymbol{P}^n$。

**定义 2.1.2**　设 $\{X_n; n \geqslant 0\}$ 为有限齐次马尔可夫链,$p_{ij}^{(n)}$ 为 $n$ 步转移概率,若存在 $n \geqslant 1$,使 $p_{ij}^{(n)} > 0$,称状态 $i$ 可达状态 $j$,记 $i \to j$。否则称为 $i$ 不达 $j$,记作 $i \nrightarrow j$。若有 $i \to j$ 和 $j \to i$ 同时成立,称 $i$ 与 $j$ 互通,记作 $i \leftrightarrow j$。

**定理 2.1.3**　若 $i \to j, j \to k$,则 $i \to k$。

**证明**　由 $i \to j, j \to k$ 必存在 $m, n$ 使

$$p_{ij}^{(n)} > 0 \qquad p_{jk}^{(m)} > 0$$

于是

$$p_{ik}^{(n+m)} \geqslant p_{ij}^{(n)} \cdot p_{jk}^{(m)} > 0$$

从而 $i \to k$。

由定理2.1.3可知,若 $i \leftrightarrow j, j \leftrightarrow k$,则 $i \leftrightarrow k$。按互通关系将状

态空间 $S$ 分类。记

$$C(i) = \{i\} \cup \{j \neq i;\ j \leftrightarrow i\}$$

分类具有以下性质：

(1) 若 $i_1 \leftrightarrow i_2$，则 $C(i_1) = C(i_2)$；

(2) 若 $i_1 \rightarrow i_2$，$i_2 \nrightarrow i_1$，则 $C(i_1) \cap C(i_2) = \varnothing$；

(3) 对于任意 $i \in S$，必在某一类中。

**定义 2.1.3** 若 $i \rightarrow j$ 时必有 $j \rightarrow i$，称 $i$ 为本质状态。若存在 $j$ 使 $i \rightarrow j$，但 $j \nrightarrow i$，称 $i$ 为非本质的。

同一类中的状态或全为本质的，或全为非本质的。本质类中的状态不可能到达别的类中，非本质类状态可能到达别的类。一旦进入本质类不可再返回。

**定义 2.1.4** $S$ 中子集 $C$ 称为闭的，若对于任意 $i \in C$，有

$$\sum_{j \in C} p_{ij} = 1$$

即 $C$ 中状态 $i$ 不可到达 $C$ 外状态。

显然本质类是闭集，非本质类不是闭集。若闭集 $C$ 中不存在真闭子集，称为最小闭集。若 $S$ 中不存在任意真闭子集，称 $S$ 是不可约的。

**例 2.1.3** 若 $\{X_n; n \geqslant 0\}$ 具有概率转移矩阵

$$\boldsymbol{P} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

显然状态集 $C = \{3,4\}$ 是闭集，而 $F = \{1,2\}$ 是非本质类。于是

$$S = C \cup F$$

一般来说，任何马尔可夫链的状态集均可分解为闭集 $C$ 和非本质类 $F$。闭集 $C$ 可以进一步分解为有限个闭集之并。

对于一个状态 $j \in S_l$，称

$$\tau_j = \min\{n \geqslant 1; \hat{X}_n = j\} \tag{2.1.3}$$

为 $\hat{X}$ 首次进入状态 $j$ 的时刻。$\{n \geqslant 1; \hat{X}_n = j\} = \varnothing$，则 $\tau_j = \infty$。令

$$f_{ij}^{(n)} = P\{\tau_j = n / \hat{X}_0 = i\} \tag{2.1.4}$$

$f_{ij}^{(n)}$ 表示从 $i$ 出发经过 $n$ 步首次到达 $j$ 的概率。于是

$$\begin{aligned} f_{ij} &= \sum_{n=1}^{\infty} f_{ij}^{(n)} = P\{\tau_j < \infty / \hat{X}_0 = i\} \\ &= P\{\bigcup_{n=1}^{\infty}\{\hat{X}_n = j\} / \hat{X}_0 = i\} \end{aligned} \tag{2.1.5}$$

是从状态 $i$ 出发迟早到达状态 $j$ 的概率。显然有

$$0 \leqslant f_{ij}^{(n)} \leqslant p_{ij}^{(n)} \leqslant f_{ij} \leqslant 1 \tag{2.1.6}$$

**定义 2.1.5**　称 $f_{ij}$ 为状态 $i$ 到状态 $j$ 的首达概率，$f_{ii} = 1$ 称状态 $i$ 为常返的。若 $f_{ii} < 1$ 称状态 $i$ 为非常返的。

对于常返状态 $i$ 有

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1 \tag{2.1.7}$$

于是 $f_{ii}^{(n)}$ 是状态 $i$ 返回时间的概率分布，从而

$$m_{ii} = \sum_{n=1}^{\infty} n f_{ii}^{(n)} \tag{2.1.8}$$

是常返状态 $i$ 的平均返回时间。

**定义 2.1.6**　对于常返状态 $i$，若 $m_{ii} < \infty$，称状态 $i$ 为正常返的。若 $m_{ii} = \infty$ 称状态 $i$ 为零常返的。

从状态 $i$ 出发无限多次到达状态 $j$ 的概率为

$$\begin{aligned} g_{ij} &= P\{\text{有无穷多个 } n \geqslant 1, \hat{X}_n = j / \hat{X}_0 = i\} \\ &= P\{\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} \{\hat{X}_n = j\} / \hat{X}_0 = i\} \end{aligned}$$

**定理 2.1.4**　状态 $i$ 为常返状态当且仅当 $g_{ii} = 1$，状态 $i$ 为非常返状态当且仅当 $g_{ii} = 0$。

**证明**　记

$$g_{ij}(m) = P\{\text{至少有 } m \text{ 个 } n \geqslant 1 \text{ 使 } X_n = j / X_0 = i\}$$

于是 $g_{ij}(1) = f_{ij}$，且 $g_{ij}(m)$ 单调下降收敛到 $g_{ij}$，由于

$$g_{ij}(m+1) = \sum_{k=1}^{\infty} P\{\tau_j = k, \text{至少有 } m \text{ 个 } n \geqslant 1 \text{ 使 } X_n = j / X_0 = i\}$$

$$= \sum_{k=1}^{\infty} f_{ij}^{(k)} g_{jj}(m) = f_{ij} \cdot g_{jj}(m)$$

特别有

$$g_{ii}(m+1) = f_{ii} g_{ii}(m) = f_{ii}^2 g_{ii}(m-1)$$

$$= (f_{ii})^m g_{ii}(1) = (f_{ii})^{(m+1)}$$

于是有

$$g_{ij}(m+1) = f_{ij}(f_{jj})^m \tag{2.1.9}$$

$$g_{ii} = \lim_{n \to \infty} g_{ii}(n) = \lim_{n \to \infty} (f_{ii})^n \tag{2.1.10}$$

由式(2.1.10)则证。

**定理 2.1.5**　若 $i$ 为常返状态，$i \to j$，则

$$g_{ji} = f_{ji} = 1$$

常返状态必为本质状态。

**证明**　由于对于任意 $m \geqslant 1$ 有

$$g_{ii} = \sum_{k=1}^{\infty} p_{ik}^{(m)} g_{ki}$$

于是对于常返状态 $i$ 有

$$1 - g_{ii} = \sum_{k=1}^{\infty} p_{ik}^{(m)} (1 - g_{ki}) = 0$$

也即对于一切 $m \geqslant 1$ 及 $k \in S$ 有

$$p_{ik}^{(m)} (1 - g_{ki}) = 0$$

若 $i \to j$，必有 $m \geqslant 1$ 使 $p_{ij}^{(m)} > 0$，于是 $g_{ji} = 1$，从而 $f_{ji} = 1$，即证 $j \to i$。

**定理 2.1.6** 若 $i$ 是常返状态，$i \to j$，则 $j$ 也是常返状态。

**证明** 设 $v_i$ 为 $i$ 在 $\{X_n\}$ 中出现的次数，于是

$$v_i = \sum_{n=1}^{\infty} \chi_{\{X_n\}}(i)$$

从而

$$E_i(v_i) = E(v_i / X_0 = i)$$

$$= \sum_{n=1}^{\infty} E(\chi_{\{X_n\}}(i)/X_0 = i) = \sum_{n=1}^{\infty} p_{ii}^{(n)}$$

于是状态 $i$ 常返的充要条件为

$$E_i(v_i) = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$$

若 $i$ 是常返状态，$i \to j$，则 $j \to i$，于是存在 $n, m \geqslant 1$ 使

$$p_{ij}^{(n)} > 0 \qquad p_{ji}^{(m)} > 0$$

从而有

$$\sum_{k=0}^{\infty} p_{jj}^{(k)} \geqslant \sum_{k=0}^{\infty} p_{jj}^{(n+m+k)} \geqslant p_{ji}^{(m)} p_{ij}^{(n)} \sum_{k=0}^{\infty} p_{ii}^{(k)} = \infty$$

从而 $j$ 也是常返状态。

由定理 2.1.6 可知，常返状态是类的性质。同一类的性质或为常返的，或为非常返的。若 $i, j$ 同属于一个常返类，则 $g_{ij} = f_{ij} = 1$，从而常返类中任意一个状态出发必须无限多次经过类中任何一个状态。与此相反非常返类中任何一个状态出发，只能有限次经过类中任意一个状态。因此非常返类是一个闭集时只能是一个无限状态集。因此对于有限的马尔可夫链非常返类必为非本质类。由于

$$g_{ij} = f_{ij} g_{jj}$$

当 $i \leftrightarrow j$ 时，$f_{ij} > 0$，于是 $g_{ij} = 0$ 当且仅当 $g_{jj} = 0$，于是 $i$ 有限次到达 $j$，当且仅当 $j$ 有限次返回 $j$。当本质类是有限状态时，$i$ 必须无限次返回 $i$，于是 $i$ 为常返状态。因此对于有限状态的马尔可夫链，本质类必为常返类，而且返回 $i$ 的平均时间是有限的，于是必为正

常返类。由以上分析可见,有限状态的马尔可夫链可分解为有限个正常返类(本质类)和一个非常返类(非本质类)。对于不可约的有限状态的马尔可夫链的状态空间是一个正常返类。

给出有限齐次马尔可夫链的转移矩阵 $\boldsymbol{P}$,即

$$\boldsymbol{P} = (p_{ij}; i, j \in S)$$

则有以下性质:

(1) 若 $C$ 是一个闭集,则对于任意 $i \in C$ 有

$$\sum_{j \in C} p_{ij} = 1$$

即

$$\boldsymbol{P}_c = (p_{ij}; i, j \in C)$$

是一个随机矩阵。正常返类(本质类)是一个闭集。

(2) 若 $F$ 是非常返类(非本质类),则对于任意 $i \in F$

$$\sum_{j \in F} p_{ij} \neq 1 \qquad \sum_{j \in S} p_{ij} = 1$$

假定 $S$ 分解为 $m$ 个正常返的闭集类和一个非常返类 $F$,则 $\boldsymbol{P}$ 可表示为

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{U} & 0 \\ \boldsymbol{V} & \boldsymbol{W} \end{pmatrix} = \begin{pmatrix} \boldsymbol{U}_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \boldsymbol{U}_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & \boldsymbol{U}_m & 0 \\ \boldsymbol{Q}_1 & \cdots & \cdots & \cdots & \boldsymbol{Q}_m & \boldsymbol{W} \end{pmatrix}$$

其中 $\boldsymbol{U}$ 与 $\boldsymbol{W}$ 是方阵,$\boldsymbol{U}_i(i \leqslant m)$ 也是方阵。$\boldsymbol{U}$ 和 $\boldsymbol{U}_i(i \leqslant m)$ 是随机矩阵。

**定义 2.1.7** 若随机矩阵 $\boldsymbol{P}$,存在 $k \geqslant 1$,使 $P^k > 0$,即对于任意 $i, j \in S$,均有 $p_{ij}^{(k)} > 0$,称 $\boldsymbol{P}$ 为本原随机矩阵。

**定理 2.1.7** 设 $\boldsymbol{P}$ 是本原随机矩阵,则存在一个平稳分布,即

存在分布$\pi$，使$\pi^{\mathrm{T}}\boldsymbol{P} = \pi^{\mathrm{T}}$，也即对于任意$j$有

$$\sum_i \pi_i p_{ij} = \pi_j > 0$$

**证明**　设$\boldsymbol{P}$为$N \times N$阶方阵，对于分布向量

$$\boldsymbol{P}^{\mathrm{T}} = (p_1, \cdots, p_N)$$

记

$$r(p) = \min_j \frac{\sum_{i=1}^{N} p_i p_{ij}}{p_j}$$

于是

$$p_j \cdot r(p) \leqslant \sum_{i=1}^{N} p_i p_{ij}$$

$$r(p) \cdot \sum_{j=1}^{N} p_j \leqslant \sum_{i=1}^{N} p_i \cdot \sum_{j=1}^{N} p_{ij} = 1$$

于是$r(p) \leqslant 1$。取

$$r = \sup\left\{r(p); \sum_{i=1}^{N} p_i = 1, p_i \geqslant 0 (i \leqslant N)\right\}$$

由于

$$A = \left\{\boldsymbol{P} = (p_1, \cdots, p_N); \sum_{i=1}^{N} p_i = 1, p_i \geqslant 0 (i \leqslant N)\right\}$$

是有界闭集，$r(p)$是$A$上有界($\leqslant 1$)的上半连续函数，因此存在$\pi \in A$，使

$$r = r(\pi) = \min_j \frac{\sum_{i=1}^{N} \pi_i p_{ij}}{\pi_j}$$

于是

$$r \pi_j \leqslant \sum_{i=1}^{N} \pi_i p_{ij}$$

即

$$r\pi^{\mathrm{T}} \leqslant \pi^{\mathrm{T}}\boldsymbol{P}$$

记

$$\boldsymbol{Z}^{\mathrm{T}} = \pi^{\mathrm{T}}\boldsymbol{P} - r\pi^{\mathrm{T}}$$

于是

$$\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{P}^{k} = (\pi^{\mathrm{T}}\boldsymbol{P}^{k})\cdot\boldsymbol{P} - r(\pi^{\mathrm{T}}\boldsymbol{P}^{k})$$

若 $\boldsymbol{Z} \neq 0$,则由 $\boldsymbol{P}^{k} > 0$,即得 $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{P}^{k} > 0^{\mathrm{T}}$,于是对于任意 $j$ 有

$$r < \frac{((\pi^{\mathrm{T}}\boldsymbol{P}^{k})\cdot\boldsymbol{P})_{j}}{(\pi^{\mathrm{T}}\cdot\boldsymbol{P}^{k})_{j}}$$

$\pi^{\mathrm{T}}\boldsymbol{P}^{k}$ 是一个分布,这与 $r$ 的定义矛盾。于是必有 $\boldsymbol{Z} = 0$,即

$$\pi^{\mathrm{T}}\boldsymbol{P} = r\pi^{\mathrm{T}}$$

由于$\pi^{\mathrm{T}}\boldsymbol{P}$ 和 $\pi^{\mathrm{T}}$ 都是分布,必有 $r = 1$,于是

$$\pi^{\mathrm{T}}\boldsymbol{P} = \pi^{\mathrm{T}}$$

即$\pi$ 是平稳分布。

**定义 2.1.8** 对于马尔可夫链的状态 $i$,正整数集$\{n \geqslant 1; p_{ii}^{(n)} > 0\}$ 的最大公因子 $d_i$ 称为状态$i$ 的周期。若 $d_i > 1$ 称状态 $i$ 为周期的;若 $d_i = 1$ 称状态 $i$ 为非周期的。

**定理 2.1.8** 若 $i \leftrightarrow j$,则 $i$ 与$j$ 有相同的周期。

**证明** 由于 $i \leftrightarrow j$,必有 $l \geqslant 1, n \geqslant 1$,使 $p_{ij}^{(l)} > 0, p_{ji}^{(n)} > 0$,于是

$$p_{ii}^{(l+n)} \geqslant p_{ij}^{(l)}\cdot p_{ji}^{(n)} > 0, \qquad p_{jj}^{(n+l)} \geqslant p_{ji}^{(n)}\cdot p_{ij}^{(l)} > 0$$

假定 $d_i$ 与$d_j$ 分别为$i$ 与$j$ 的周期,则$(l+n)$同时可被 $d_i$ 和$d_j$ 整除。若 $p_{ii}^{(m)} > 0$,则

$$p_{jj}^{(l+m+n)} \geqslant p_{ji}^{(n)}\cdot p_{ii}^{(m)}\cdot p_{ij}^{(l)} > 0$$

$(l+m+n)$可被 $d_j$ 整除,从而 $m$ 也可被$d_j$ 整除,从而 $d_i$ 可被$d_j$ 整除。同理 $d_j$ 可被$d_i$ 整除,于是 $d_i = d_j$。

**定理 2.1.9** 若马尔可夫链是不可约的,即所有状态之间互

通，且存在 $j \leqslant N$，使 $p_{jj} > 0$，则此链是非周期的，其转移矩阵为本原随机矩阵，存在平稳分布。

**证明** 由于 $p_{jj} > 0$，则 $d_j = 1$。由定理2.1.8知，所有状态周期均为1。由于对于任意状态 $i, i \leftrightarrow j$，存在 $n_{ij}$ 和 $n_{ji}$ 使 $p_{ij}^{(n_{ij})} > 0$，$p_{ji}^{(n_{ji})} > 0$，于是

$$p_{ij}^{(n_{ij}+n_{ji})} > 0$$

同时有

$$p_{ij}^{(n_{ij}+n_{ji}+n)} \geqslant p_{ij}^{(n_{ij})} p_{jj} p_{jj} \cdots p_{jj} p_{ji}^{n_{ji}} > 0$$

取 $N_0 = \max\{n_{ij} + n_{ji}; i, j \leqslant N\}$，则 $k \geqslant N_0$ 时 $p_{ij}^{(k)} > 0(i, j \leqslant N)$，于是 $p_{ij}^{(k)} > 0(k \geqslant N_0)$，$\boldsymbol{P}$ 为本原随机矩阵。根据定理2.1.7马尔可夫链存在平稳分布。

**定理 2.1.10** 非周期不可约有限状态的马尔可夫链的转移随机矩阵 $\boldsymbol{P}$ 存在平稳分布。

**证明** 设 $\boldsymbol{P}$ 为转移矩阵，不可约的，则所有状态是正常返的，而且是互通的。若 $\boldsymbol{P}$ 为非周期的，则对于任意 $i \leqslant N$，$\{n \geqslant 1; p_{ii}^{(n)} > 0\}$ 的最大公因子是1。于是存在 $n_i$，当 $n \geqslant n_i$ 时 $p_{ii}^{(n)} > 0$。取 $M = \max\limits_{i \leqslant N}\{n_i\}$，则 $n \geqslant M$ 时，对于任意 $i \leqslant N$ 有 $p_{ii}^{(n)} > 0$。由于 $i \leftrightarrow j$，则存在 $n_{ij}$ 使 $p_{ij}^{(n_{ij})} > 0$，于是 $n \geqslant M$ 时，$p_{ij}^{(n_{ij}+n)} > 0$。取

$$K = M + \max_{ij}\{n_{ij}\}$$

则对于任意 $i, j \leqslant N$，只要 $k \geqslant K$，总有 $p_{ij}^{(n)} > 0$。于是 $\boldsymbol{P}$ 是本原随机矩阵。由定理2.1.7知，$\boldsymbol{P}$ 存在平稳分布。

**定理 2.1.11** 若齐次马尔可夫链转移矩阵 $\boldsymbol{P}$，对于任何初始分布 $q$，有极限分布

$$\lim_{n \to \infty} q^{\mathrm{T}} \boldsymbol{P}^n = \pi^{\mathrm{T}}$$

则极限分布 $\pi$ 是 $\boldsymbol{P}$ 的平稳分布。

**证明** 由于

$$q^{\mathrm{T}}\boldsymbol{P}^{n+1} = q^{\mathrm{T}}\boldsymbol{P}^{n}\cdot\boldsymbol{P}$$

则

$$\pi^{\mathrm{T}} = \lim_{n\to\infty} q^{\mathrm{T}}\boldsymbol{P}^{n+1} = (\lim_{n\to\infty}(q^{\mathrm{T}}\boldsymbol{P}^{n}))\cdot\boldsymbol{P} = \pi^{\mathrm{T}}\cdot\boldsymbol{P}$$

得证。

由于 $j$ 为非常返状态时，$\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$，于是 $\lim_{n\to\infty} p_{jj}^{(n)} = 0$。若 $j$ 为常返状态时且是非周期的，则有

$$\lim_{n\to\infty} p_{jj}^{(n)} = \frac{1}{m_{jj}}$$

于是对零常返状态 $j$ 有 $\lim_{n\to\infty} p_{jj}^{(n)} = 0$。由于

$$\begin{aligned} p_{jk}^{(n)} &= \sum_{m=1}^{n} f_{jk}^{(m)} p_{kk}^{(n-m)} \\ &\leqslant \sum_{m=1}^{N} f_{jk}^{(m)} p_{kk}^{(n-m)} + \sum_{m=N+1}^{n} f_{jk}^{(m)} \end{aligned}$$

先令 $n\to\infty$，再令 $N\to\infty$，则对 $k$ 为非常返状态和零常返状态有

$$\lim_{n\to\infty} p_{jk}^{(n)} = 0$$

于是对于有限状态的马尔可夫链不可能存在零常返状态。对于非周期的正常返的状态 $k$ 有

$$\sum_{m=1}^{N} f_{jk}^{(m)} p_{kk}^{(n-m)} \leqslant p_{jk}^{(n)} \leqslant \sum_{m=N+1}^{\infty} f_{jk}^{(m)} + \sum_{m=1}^{N} f_{jk}^{(m)} p_{kk}^{(n-m)}$$

首先令 $n\to\infty$，再令 $N\to\infty$，由 $\lim_{n\to\infty} p_{kk}^{(n)} = \frac{1}{m_{kk}}$ 及 $\sum_{m=1}^{\infty} f_{jk}^{(m)} = 1$，则得

$$\lim_{n\to\infty} p_{jk}^{(n)} = \frac{1}{m_{kk}}$$

**定理 2.1.12** 对于不可约非周期的马尔可夫链有极限分布

$$\lim_{n\to\infty} P\{X_n = j\} = \pi_j = \frac{1}{m_{jj}}$$

此极限分布与初始分布无关。

**证明**　由于

$$\lim_{n\to\infty}P\{X_n=j\}=\lim_{n\to\infty}\sum_k q_k p_{kj}^{(n)}$$
$$=\sum_k q_k\lim_{n\to\infty}p_{kj}=\frac{1}{m_{jj}}\sum_k q_k=\frac{1}{m_{jj}}$$

则证。

**定理 2.1.13**　假定马尔可夫链仅有一个正常返类 $C$，则当 $j\in C$ 时

$$\lim_{n\to\infty}P\{X_n=j\}=\pi_j$$

且 $j\notin C$ 时

$$\lim_{n\to\infty}P\{X_n=j\}=0$$

**证明**　利用定理 2.1.12 则证。

## 2-2　标准遗传算法的马氏链模型

假设 $S=\{0,1\}^l$ 为个体空间，$S^N$ 为种群空间，$S^2$ 为母体空间。若 $X_i\in S(i\leqslant N)$ 是 $N$ 个个体，则

$$\vec{X}=(X_1,\cdots,X_N)$$

是一个种群。用 $\vec{X}(0)$ 表示初始种群，$\vec{X}(n)$ 表示第 $n$ 代种群。用 $f:S\to R^+$ 表示适应值函数，则标准遗传算法的过程为：

(1) 对于 $n=0$ 给出初始种群 $\vec{X}(0)$；

(2) 对于第 $n$ 步在 $\vec{X}(n)$ 中按以下分布独立地选择 $N$ 个母体：

$$P\{T_s(\vec{X}(n))=(Y_1^{(k)},Y_2^{(k)})\}=\begin{cases}f(Y_1^{(k)})\cdot f(Y_2^{(k)})/\left(\sum\limits_{X\in\vec{X}(n)}f(X)\right)^2, & (Y_1^{(k)},Y_2^{(k)})\in\vec{X}(n)\\0, & \text{否则}\end{cases}\tag{2.2.1}$$

其中 $k = 1 \sim N$。

(3) 对(2)中选择的 $N$ 个母体进行单点随机杂交,即

$$P\{T_c(Y_1^{(k)}, Y_2^{(k)}) = X_k'(n+1)\}$$

$$= \begin{cases} kp_c/l, & X_k'(n+1) \neq Y_1^{(k)} \\ & \text{且 } X_k'(n+1) = AY_1^{(k)} + (I-A)Y_2^{(k)} \\ (1-p_c) + \dfrac{kp_c}{l}, & X_k'(n+1) = Y_1^{(k)} \\ 0, & \text{其它} \end{cases}$$

(2.2.2)

其中 $p_c$ 满足 $0 < p_c \leqslant 1$,$p_c$ 是杂交概率。$A$ 是一个对角线矩阵,前 $r$ 个对角元素为 1 而其它元素为 0。$r$ 称为单点杂交的杂交点。

(4) 设变异概率为 $p_m$,通过

$$\vec{X}'(n+1) = (X_1'(n+1), \cdots, X_N'(n+1))$$

产生

$$\vec{X}(n+1) = (X_1(n+1), \cdots, X_N(n+1))$$

满足概率:

$$P\{T_m(X_k'(n+1)) = X_k(n+1)\}$$

$$= p_m^{d(X_k'(n+1), X_k(n+1))}(1-p_m)^{l-d(X_k'(n+1), X_k(n+1))}(k \leqslant N)$$

(2.2.3)

(5) 如果满足停止准则,则停止,否则转向(2)。

记

$$T = T_m \circ T_c \circ T_s$$

则标准遗传算法为:

$$\vec{X}(n) = T(\vec{X}(n-1)) = T_m \circ T_c \circ T_s(\vec{X}(n-1))$$

(2.2.4)

从 $\vec{X}(0)$ 出发,即得到种群序列 $\{\vec{X}(n); n \geqslant 0\}$。

**定理 2.2.1** 标准遗传算法的种群序列 $\{\vec{X}(n); n \geqslant 0\}$ 是有

限齐次不可约非周期的马尔可夫链。

**证明**　由于 $S=\{0,1\}^l$ 中有 $2^l$ 个个体，则种群空间 $S^N$ 中有 $2^{Nl}$ 个个体，$S^N$ 为种群序列的状态空间，是有限的。由于

$$\vec{X}(n+1)=T(\vec{X}(n))=T_m\circ T_c\circ T_s(\vec{X}(n))$$

其中 $T_m,T_c,T_s$ 均与 $n$ 无关，因此 $\vec{X}(n+1)$ 仅与 $\vec{X}(n)$ 有关，即 $\{\vec{X}(n);n\geqslant 0\}$ 是有限状态的马尔可夫链。由于

$$\begin{aligned}
&P\{(T(\vec{X}(n))_k=X_k(n+1)\}\\
&=\sum_{X_k'(n+1)\in S}\ \sum_{(Y_1^{(k)}(n),Y_2^{(k)}(n))\in S^2}P\{T_s(\vec{X}(n)\}=(Y_1^{(k)}(n),\\
&\quad Y_2^{(k)}(n))\}\cdot P\{T_c(Y_1^{(k)}(n),Y_2^{(k)}(n))=X_k'(n+1)\}\\
&\quad\cdot P\{T_m(X_k'(n+1)=X_k(n+1)\}
\end{aligned}\tag{2.2.5}$$

及对于任意 $\vec{X}(n)\in S^N$，均存在 $(Y_1^{(k)}(n),Y_2^{(k)}(n))\in S^N$ 及 $X_k'(n+1)$ 使得

$$P\{T_s(\vec{X}(n))=(Y_1^{(k)}(n),Y_2^{(k)}(n))\}>0\tag{2.2.6}$$

$$P\{T_c(Y_1^{(k)}(n),Y_2^{(k)}(n))=X_k'(n+1)\}>0\tag{2.2.7}$$

对于任意 $X_k'(n+1)$ 及 $X_k(n+1)$ 有

$$P\{T_m(X_k'(n+1))=X_k(n+1)\}>0\tag{2.2.8}$$

于是

$$P\{(T(\vec{X}(n)))_k=X_k(n+1)\}>0\tag{2.2.9}$$

由式(2.2.1)，(2.2.2) 及(2.2.3) 知式(2.2.6)，(2.2.7) 及(2.2.8) 与 $n$ 无关，从而式(2.2.9) 也与 $n$ 无关。从而

$$\begin{aligned}
&P\{T(\vec{X}(n))=\vec{X}(n+1)\}\\
&=\prod_{k=1}^{N}P\{(T(\vec{X}(n))_k=X_k(n+1)\}>0
\end{aligned}$$

且与 $n$ 无关。

由以上证明可见，标准遗传算法的种群序列 $\{\vec{X}(n);n\geqslant 0\}$ 是 $S^N$ 上的齐次不可约非周期的马尔可夫链。

**定理 2.2.2** 标准遗传算法的种群马尔可夫链序列$\{\vec{X}(n);n \geqslant 0\}$存在极限分布。

**证明** 由定理2.2.1,标准遗传算法的种群序列$\{\vec{X}(n);n \geqslant 0\}$是有限状态不可约的非周期的齐次马尔可夫链,根据定理2.1.12知

$$\lim_{n \to \infty} P\{\vec{X}(n) = \vec{Y}\}$$

$$= \lim_{n \to \infty} \sum_{\vec{X}(0) \in S^N} P\{\vec{X}(n) = \vec{Y}/\vec{X}(0)\} P\{\vec{X}(0)\} = \pi(\vec{Y})$$

则$\pi(\vec{Y})$是$S^N$上的分布,且

$$\sum_{\vec{Y} \in S^N} \pi(\vec{Y}) P\{\vec{X}(n) = \vec{Z}/\vec{X}(0) = \vec{Y}\} = \pi(\vec{Z})$$

定理2.2.2说明标准遗传算法的种群马尔可夫链是分布收敛的,它不依赖于初始种群的选择,即

$$\pi(\vec{Z}) = \lim_{n \to \infty} P\{\vec{X}(n) = \vec{Z}/\vec{X}(0) = \vec{Y}\} \qquad (\vec{Z} \in S^N)$$

假定$f$是$S$上的适应值函数,记

$$M = \{X_i \in S, 使\ f(X_i) = \max_{X \in S} f(X)\}$$

若$\vec{Y} \cap M = \varnothing$,则

$$\lim_{n \to \infty} P\{\vec{X}(n) \cap M \neq \varnothing / \vec{X}(0) = \vec{Z}\} \leqslant 1 - \pi(\vec{Y}) < 1$$

即不能保证,当$n \to \infty$时$\vec{X}(n) \cap M \neq \varnothing$。因此标准遗传算法不能保证得到全局最优解。

但是,由于标准遗传算法的状态空间是正常返的,它以概率1保证$\{\vec{X}(n);n \geqslant 0\}$可以达到$S^N$中任何状态无限多次。特别对于$\vec{X} \in M$,也可以到达无限多次,从实用的角度来讲也是有效的。

在定理2.2.1和定理2.2.2中使用单点杂交,事实上适合于任何杂交方式。

下面我们进一步研究标准遗传算法在种群转移过程中遗传因子的变化规律。记

$$P_i\{\vec{X}, y_j\} = P\{x_{ij}(n+1) = y_j / \vec{X}(n) = \vec{X}\} (y_j = 0, 1)$$

其中

$$\vec{X}=(X_1,\cdots,X_N)\in S^N$$

$$X_i=(x_{i1},\cdots,x_{il})\in S \qquad (i\leqslant N)$$

称 $P_i\{\vec{X},y_j\}$ 为给定 $\vec{X}$ 时,后代个体的一维边际分布,令

$$I_1^j=\{i\leqslant N;\ x_{ij}=1\}$$

$$I_0^j=\{i\leqslant N;\ x_{ij}=0\}$$

$$f_1^j(\vec{X})=\sum_{i\in I_1^j}f(X_i)=\sum_{x_{ij}=1}f(X_i)$$

$$f_0^j(\vec{X})=\sum_{i\in I_0^j}f(X_i)=\sum_{x_{ij}=0}f(X_i)$$

$$\bar{f}(\vec{X})=\sum_{i=1}^{N}f(X_i)=f_0^j(\vec{X})+f_1^j(\vec{X})$$

**定理2.2.3**　设$\{\vec{X}(n);n\geqslant 0\}$ 为采用单点杂交算子的标准遗传算法的种群序列。给定 $\vec{X}(n)=\vec{X}$,则下代个体 $X_i(n+1)$ 的第 $j$ 个分量 $x_{ij}(n+1)$ 服从参数为 $p_j(\vec{X},1)$ 的 0-1 分布,其中

$$p_j(\vec{X},1)=P\{x_{ij}(n+1)=1/\vec{X}(n)=\vec{X}\}$$

$$=\frac{f_1^j(\vec{X})}{\bar{f}(\vec{X})}+\left[1-2\frac{f_1^j(\vec{X})}{\bar{f}(\vec{X})}\right]p_m \qquad (2.2.10)$$

**证明**　令

$$Y=(y_1,\cdots,y_l)=T_c(T_s(\vec{X}))$$

则

$$P\{x_{ij}(n+1)=1/\vec{X}(n)=\vec{X}\}$$

$$=P\{y_j=1/\vec{X}(n)=\vec{X}\}\cdot(1-p_m)$$

$$+P\{y_j=0/\vec{X}(n)=\vec{X}\}\cdot p_m \qquad (2.2.11)$$

而

$$P\{y_j=1/\vec{X}(n)=\vec{X}\}$$

$$=P\{Y=T_c(T_s(\vec{X}));y_j=1\}$$

$$= \sum_{u \leqslant N} \sum_{v \leqslant N} P\{T_c(X_u, X_v) = Y; y_j = 1\} \cdot P\{T_s(\vec{X}) = (X_u, X_v)\}$$

记

$$A_1 = P_1^j \times P_1^j, \qquad A_2 = P_1^j \times P_0^j$$

$$A_3 = P_0^j \times P_1^j, \qquad A_4 = P_0^j \times P_0^j$$

于是 $N \times N = A_1 \cup A_2 \cup A_3 \cup A_4$。记

$$B_i = \sum_{(u,v) \in A_i} P\{T_c(X_u, X_v) = Y; y_j = 1\} \cdot P\{T_s(\vec{X}) = (X_u, X_v)\} \quad (i \leqslant 4)$$

于是

$$P\{y_i = 1 / \vec{X}(n) = \vec{X}\} = B_1 + B_2 + B_3 + B_4$$

当$(u,v) \in A_4$时,$T_c(X_u, X_v) = Y$中$y_i = 0$,于是$B_4 = 0$。

当$(u,v) \in A_1$时,$T_c(X_u, X_v) = Y$中$y_i = 1$,于是

$$B_1 = \sum_{(u,v) \in A_1} P\{T_s(\vec{X}) = (X_u, X_v)\}$$

$$= \sum_{(u,v) \in A_1} \frac{f(X_u) f(X_v)}{(\bar{f}(\vec{X}))^2}$$

当$(u,v) \in A_3$时,由于$u \in P_0^j, v \in P_1^j$,即$x_{uj} = 0, x_{vj} = 1$,于是杂交必须在$j$分量之前进行才能使$y_j = 1$,于是

$$B_3 = \sum_{(u,v) \in A_3} \frac{j p_c}{l} \frac{f(X_u) f(X_v)}{(\bar{f}(\vec{X}))^2}$$

当$(u,v) \in A_2$时,由于$u \in P_1^j, v \in P_0^j$,即$x_{uj} = 1, x_{vj} = 0$,于是要使$y_j = 1$,必须满足下面两种情形之一:

(a) 杂交在$j$分量之后进行,这时不管是否杂交均有$y_j = 1$,于是

$$P\{T_c(X_u, X_v) = y, y_i = 1\} = \frac{l - j}{l}$$

(b) 杂交在 $j$ 分量之前进行，但事实上未杂交，于是

$$P\{T_c(X_u,X_v)=y,y_i=1\}=\frac{j}{l}(1-p_c)$$

从而

$$\begin{aligned}B_2 &= \sum_{(u,v)\in A_2}\left(\frac{l-j}{l}+\frac{j}{l}(1-p_c)\right)\\&\quad\cdot P\{T_s(\vec{X})=(X_u,X_v)\}\\&=\sum_{(u,v)\in A_2}\left(1-\frac{j}{l}p_c\right)\frac{f(X_u)f(X_v)}{(\bar{f}(\vec{X}))^2}\end{aligned}$$

综合上述结果即得

$$\begin{aligned}P\{y_i &= 1/\vec{X}(n)=\vec{X}\}=B_1+B_2+B_3\\&=\frac{\left(\sum_{u\in I_1'}f(X_u)\right)^2}{(\bar{f}(\vec{X}))^2}+\sum_{(u,v)\in A_2}\frac{f(X_u)f(X_v)}{(\bar{f}(\vec{X}))^2}\\&=\left[\frac{f_1(\vec{X})}{(\bar{f}(\vec{X}))}\right]^2+\frac{f_1(\vec{X})}{(\bar{f}(\vec{X}))}\left[1-\frac{f_1(\vec{X})}{\bar{f}(\vec{X})}\right]\\&=\frac{f_1(\vec{X})}{\bar{f}(\vec{X})}\end{aligned}$$

由式(2.2.11)即得式(2.2.10)。

**定理 2.2.4**　设 $\{\vec{X}(n);n\geqslant 0\}$ 是采用均匀杂交算子的标准遗传算法的种群马尔可夫链，则

$$\begin{aligned}P\{y_j &= 1/\vec{X}(n)=\vec{X}\}\\&=\frac{f_1(\vec{X})}{\bar{f}(\vec{X})}+\left[1-2\frac{f_1(\vec{X})}{(\bar{f}(\vec{X}))}\right]p_m\end{aligned}\tag{2.2.12}$$

**证明**　利用定理 2.2.3 中记号有

$$P\{y_i=1/\vec{X}(n)=\vec{X}\}$$

$$= \sum_{(u,v)\in A_1} \frac{f(X_u)f(X_v)}{(\bar{f}(\vec{X}))^2} + \sum_{(u,v)\in A_2} \frac{f(X_u)f(X_v)}{(\bar{f}(\vec{X}))^2}(1-p_c)$$

$$+ \sum_{(u,v)\in A_3} \frac{f(X_u)f(X_v)}{(\bar{f}(\vec{X}))^2} p_c = \frac{\rho_1(\vec{X})}{\bar{f}(\vec{X})}$$

则证。

**推论 2.2.1** 设$\{\vec{X}(n); n \geqslant 0\}$为标准遗传算法的种群马尔可夫链，而

$$\mathscr{L}(\vec{X}) = \mathscr{L}[(i_k, a_k); 0 \leqslant k \leqslant K]$$

为含 $\vec{X}$ 的最小模式，对于 $a_k = 1 (k \leqslant K)$ 时有

$$P\{y_{i_k} = 1 / \vec{X}(n) = \vec{X}\} = 1 - p_m$$

当 $a_k = 0 (k \leqslant K)$ 时有

$$P\{y_{i_k} = 1 / \vec{X}(n) = \vec{X}\} = p_m$$

**证明** 由于 $a_k = 1$ 时 $f_1^{\dagger}(\vec{X}) = \bar{f}(\vec{X})$，$a_k = 0$ 时 $f_1^{\dagger}(\vec{X}) = 0$，则证。

下面研究 $p_m = 0$ 情况下的标准遗传算法。

**定理 2.2.5** 假定$\{\vec{X}(n); n \geqslant 1\}$是标准遗传算法的种群马尔可夫链，$p_m = 0$，$\vec{X}(0) = \vec{X}_0$。有以下结论：

(1) 对于任意 $Y \in \mathscr{L}(\vec{X}_0)$，存在 $n \geqslant 0$，使得

$$P\{Y \in \vec{X}(n) / \vec{X}(0) = \vec{X}_0\} > 0 \tag{2.2.13}$$

(2) 对于任意 $Y \notin \mathscr{L}(\vec{X}_0)$ 和 $n \geqslant 0$ 有

$$P\{Y \in \vec{X}(n) / \vec{X}(0) = \vec{X}_0\} = 0 \tag{2.2.14}$$

其中 $\mathscr{L}(\vec{X}_0)$ 为包含 $\vec{X}_0$ 的最小模式。

**证明** (1) 记 $\boldsymbol{Y} = (y_1, \cdots, y_l)^{\mathrm{T}} \in \mathscr{L}(\vec{X}_0)$ 和

$$\vec{X}_0 = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1l} \\ x_{21} & x_{22} & \cdots & x_{2l} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nl} \end{pmatrix}$$

用 $\lambda(\vec{X}_0)$ 和 $\beta(\vec{X}_0)$ 分别表示 $\vec{X}_0$ 的多样度与成熟度。我们可以假设

$$\mathscr{L}(\vec{X}_0) = \mathscr{L}((i,a_i);\ i \leqslant \lambda(\vec{X}_0))$$
$$= (a_1, a_2, \cdots, a_{\lambda(\vec{X}_0)}, *, \cdots, *)$$

于是对于任意 $j \leqslant \lambda(\vec{X}_0)$ 有

$$x_{ij} = y_j \qquad (i \leqslant N)$$

显然对于任意 $j \geqslant \beta(\vec{X}_0) + 1$ 及 $y_j$，存在 $\vec{X}_0$ 中的个体 $x_{ij}$ 使得 $x_{i_j j} = y_j$。

下面用归纳法证明(1)。设 $\lambda(\vec{X}_0) = l - 2$，也即有 $\beta(\vec{X}_0) = 2$。这时对于任意 $i \leqslant N, j \leqslant l - 2$ 有 $x_{ij} = y_j$。同时存在 $1 \leqslant p, q \leqslant N$，使得 $X_p(0), X_q(0)$ 是 $\vec{X}_0$ 的个体，且

$$x_{p(l-1)} = y_{l-1}, \qquad x_{ql} = y_l$$

令

$$\vec{Z} = (Z_1, \cdots, Z_N)$$

对于任何 $i \neq p(i \leqslant N)$ 取 $Z_i = X_i(0)$，$Z_p$ 利用 $X_p(0)$ 和 $X_q(0)$ 在第 $l$ 个分量上杂交产生，即

$$Z_p = (x_{p1}, \cdots, x_{p(l-1)}, x_{ql}) = Y$$

利用选择算子与杂交算子的性质即得

$$P\{Y \in \vec{X}(1)/\vec{X}(0) = X_0\} \geqslant P\{\vec{X}(1) = Z/\vec{X}(0) = \vec{X}_0\}$$
$$\geqslant \left\{\prod_{i \neq p} \frac{f(X_i(0))^2}{\left(\sum_{j=1}^{N} f(X_j(0))^2\right)}\right\} \cdot \left\{\frac{p_c}{l} \frac{f(X_p(0)) \cdot f(X_q(0))}{\left(\sum_{j=1}^{N} f(X_j(0))\right)^2}\right\}$$
$$> 0$$

对于 $\beta(\vec{X}_0) = 2$ 的情况证明了式(2.2.10)。

现在假定式(2.2.13)对于 $\beta(\vec{X}_0) = m - 1$ 是正确的，考虑 $\beta(\vec{X}_0) = m$ 的情形。那么我们有

$$x_{ij} = y_j \qquad (i \leqslant N, j \leqslant l - m)$$

以及存在 $i_0(1 \leqslant i_0 \leqslant N)$ 使得 $x_{i_0 l} = y_l$。取

$$\vec{Z}(1) = (Z_1, \cdots, Z_N)^{\mathrm{T}}$$

其中

$$Z_i = (x_{i1}, \cdots, x_{i(l-m)}, \cdots, x_{i(l-1)}, y_l)$$

于是 $Y = Z_{i_0} \in \mathscr{L}(\vec{Z}(1)), \beta(\vec{Z}(1)) = m - 1, \lambda(\vec{Z}(1)) = l - m + 1$。由于 $\mathscr{L}(\vec{Z}(1)) \subset \mathscr{L}(\vec{X}(0))$，于是由归纳假定及马尔可夫链的齐次性得到

$$P\{Y \in \vec{X}(n) / \vec{X}(1) = \vec{Z}(1)\} > 0$$

进一步利用遗传算子的定义有

$$\begin{aligned} & P\{\vec{X}(1) = \vec{Z}(1) / \vec{X}(0) = \vec{X}_0\} \\ & \quad = \left\{\prod_{i \neq i_0} \frac{p_c}{l} \frac{f(X_i) f(X_{i_0})}{\left(\sum_{j=1}^{N} f(X_j)\right)^2}\right\} \cdot \left\{\frac{f(X_{i_0})^2}{\left(\sum_{j=1}^{N} f(X_j)\right)^2}\right\} > 0 \end{aligned}$$

于是

$$\begin{aligned} & P\{Y \in \vec{X}(n) / \vec{X}(0) = \vec{X}_0\} \\ & \quad \geqslant P\{Y \in \vec{X}(n) / \vec{X}(1) = \vec{Z}(1)\} \\ & \quad\quad \cdot P\{\vec{X}(1) = \vec{Z}(1) / \vec{X}(0) = \vec{X}_0\} > 0 \end{aligned}$$

(2) 由假定 $Y \notin \mathscr{L}(\vec{X}_0)$，则

$$P\{Y \in \mathscr{L}(\vec{X}_0) / \vec{X}(0) = \vec{X}_0\} = 0$$

设 $\mathscr{L}(\vec{X}_0) = \mathscr{L}[a_{i_1}, \cdots, a_{i_k}], k = \lambda(\vec{X}_0)$，对于选择算子与杂交算子不可能改变 $i_1, \cdots, i_k$ 的基因分量 $a_{i_1}, \cdots, a_{i_k}$，于是对于任意 $Y \in \mathscr{L}(\vec{X}_0)$ 有

$$P\{\vec{X}(k+1) \subset \mathscr{L}(\vec{X}_0) / \vec{X}(k) = Y\} = 1$$

由于 $\vec{X}_0 \in \mathscr{L}(\vec{X}_0)$，于是

$$P\{\vec{X}(n) \subset \mathscr{L}(\vec{X}_0) / \vec{X}(0) = \vec{X}_0\} = 1$$

联合以上证明即得，对于 $Y \notin \mathscr{G}(\vec{X}_0)$ 有

$$P\{Y \in \vec{X}(n)/\vec{X}(0) = \vec{X}_0\}$$
$$\leqslant P\{Y \in \mathscr{G}(\vec{X}_0), \vec{X}(n) \subset \mathscr{G}(\vec{X}_0)/\vec{X}(0) = \vec{X}_0\}$$
$$= P\{Y \in \mathscr{G}(\vec{X}_0)/\vec{X}(0) = \vec{X}_0\} = 0$$

则证。

定理2.2.5指出了在没有变异的情况下，标准遗传算法的搜索范围为 $2^{\lambda(\vec{X}_0)}$ 个个体。多样度 $\lambda(\vec{X}_0)$ 表征着搜索空间的大小，$\lambda(\vec{X}_0)$ 越大搜索空间越大。特别当 $\lambda(\vec{X}_0)$ 为0时没有什么搜索能力。

**定理2.2.6**　设 $\{\vec{X}(n); n \geqslant 0\}$ 是标准遗传算法的种群马尔可夫链，$p_m = 0$，$H$ 是齐次种群全体，即

$$H = \{(X, \cdots, X); X \in S\}$$

则对于任意 $n \geqslant 1$ 有

$$P\{\vec{X}(n) \in H/\vec{X}(0) \in H\} = 1$$

**证明**　对于任意 $\vec{X} \in H$，$\mathscr{G}(\vec{X}) = \vec{X}$，因而 $\vec{Y} \notin H$ 时，由定理2.2.3，对于任意 $n \geqslant 1$ 必有

$$P\{\vec{X}(n) = \vec{Y}/\vec{X}(0) = \vec{X}\} = 0$$

于是

$$P\{\vec{X}(n) \notin H/\vec{X}(0) \in H\}$$
$$= \sum_{\vec{Y} \notin H} P\{\vec{X}(n) = \vec{Y}/\vec{X}(0) \in H\}$$
$$= \sum_{\vec{Y} \notin H} \frac{P\{\vec{X}(n) = \vec{Y}, \vec{X}(0) \in H\}}{P\{\vec{X}(0) \in H\}}$$
$$= \sum_{\vec{Y} \notin H} \sum_{\vec{X} \in H} \frac{P\{\vec{X}(n) = \vec{Y}/\vec{X}(0) = \vec{X}\} P\{\vec{X}(0) = \vec{X}\}}{P\{\vec{X}(0) \notin H\}}$$
$$= 0$$

从而

$$P\{\vec{X}(n) \in H / \vec{X}(0) \in H\} = 1$$

**定理 2.2.7** 设$\{\vec{X}(n); n \geqslant 0\}$是标准遗传算法的马尔可夫链，$p_m = 0$，$H$为齐次种群，则有

(1) $\{\vec{X}(n); n \geqslant 0\}$以概率1收敛到$H$，即

$$\lim_{n \to \infty} P\{\vec{X}(n) \in H\} = 1$$

(2) $\{\beta(\vec{X}(n)); n \geqslant 1\}$以概率1单调不增，即

$$P\{\beta(\vec{X}(n+1)) \leqslant \beta(\vec{X}(n))\} = 1$$

(3) $\{\beta(\vec{X}(n)); n \geqslant 1\}$以正概率单调减，即

$$P\{\beta(\vec{X}(n+1)) < \beta(\vec{X}(n))\} > 0$$

(4) $\{\beta(\vec{X}(n)); n \geqslant 1\}$几乎处处收敛到0，即

$$P\{\lim_{n \to \infty} \beta(\vec{X}(n)) = 0\} = 1$$

**证明** (1) 如果$\vec{X}(0) \in H$，由定理2.2.6则证。下面假定$\vec{X}(0) = (X_1, \cdots, X_N) \notin H$。记

$$P\{\vec{X}, \vec{Y}\} = P\{\vec{X}(n+1) = \vec{Y} / \vec{X}(n) = \vec{X}\}$$

$$P\{\vec{X}, H\} = \sum_{\vec{Y} \in H} P\{\vec{X}, \vec{Y}\}$$

则

$$P\{\vec{X}, H\} \geqslant \sum_{i=1}^{N} P\{\vec{X}(n+1) = (X_i, \cdots, X_i) / \vec{X}(n) = \vec{X}\}$$

$$= \sum_{i=1}^{N} \left[ \frac{f(X_i)^2}{\left( \sum_{j=1}^{N} f(X_j) \right)^2} \right]^N > 0$$

记

$$a = \min\{P(Y, H);\ Y \notin H\}$$

则$0 < a < 1$。对于

$$T = \min\{k \geqslant 1; \vec{X}(k) \in H\}$$

和$k \geqslant 1$有

$$P\{T=k\} = \sum_{\vec{Y}_1,\cdots,\vec{Y}_{k-1}\notin H} P\{\vec{X},\vec{Y}_1\}P\{\vec{Y}_1,\vec{Y}_2\}\cdots P\{\vec{Y}_{k-1},H\} \leqslant (1-a)^{k-1}$$

于是

$$E(T) = \sum_{k=1}^{\infty} k\cdot\{T=k\} \leqslant \sum_{k=1}^{\infty} k(1-a)^{k-1} = \frac{1}{a^2} < \infty$$

从而 $P\{T<\infty\}=1$,即得

$$\lim_{k\to\infty} P\{\vec{X}(k)\in H\} = P\{T<\infty\} = 1$$

(2) 由于 $\mathscr{L}(\vec{X})$ 是含 $\vec{X}$ 的最小模式,则

$$\begin{aligned}
&P\{\beta(\vec{X}(k+1))\leqslant\beta(\vec{X}(k))\}\\
&=\sum_{\vec{X}\in S^N} P\{\lambda(\vec{X}(k+1))\leqslant\beta(\vec{X}(k))/\vec{X}(k)=\vec{X}\}\\
&\quad\cdot P\{\vec{X}(k)=\vec{X}\}\\
&\geqslant\sum_{\vec{X}\in S^N} P\{\vec{X}(k+1)\in\prod_{i=1}^{N}\mathscr{L}(\vec{X})/\vec{X}(k)=\vec{X}\}\\
&\quad\cdot P\{\vec{X}(k)=\vec{X}\}=1
\end{aligned}$$

(3) 对于 $\vec{X}\in S^N$ 有

$$\begin{aligned}
&P\{\beta(\vec{X}(k+1))<\beta(\vec{X})/\vec{X}(k)=\vec{X}\}\\
&=P\{\beta(\vec{X}(k+1))\leqslant\beta(\vec{X})-1/\vec{X}(k)=\vec{X}\}\\
&\geqslant P\{\beta(\vec{X}(k+1))=0/\vec{X}(k)=\vec{X}\}\\
&\geqslant\Big((f(X_1))^2/\Big(\sum_{i=1}^{N}f(X_i)\Big)^2\Big)^N>0
\end{aligned}$$

类似于(2)可证

$$P\{\beta(\vec{X}(k+1))<\beta(\vec{X}(k))\}>0$$

(4) 由(1)及

$$P\{\lim_{k\to\infty}\beta(\vec{X}(k))=0\}\geqslant P\{\lim_{k\to\infty}\vec{X}(k)\in H\}=1$$

则证。

## 2-3 改进遗传算法的马氏链模型

从定理 2.2.5 可以看到，在标准遗传算法中杂交算子有能力搜索包含当前种群的极小模式中的所有个体，但是它也仅限于此。因此，只有当种群的多样度越小时，包含种群的定义长度越大，从而包含当前种群的极小模式包含更多的可行解种群。当前种群的多样度越大，杂交算子的搜索能力越强。如果初始种群的极小模式包含全局最优解，则标准遗传算法有可能搜索到全局最优解。同时从上节的定理 2.2.6 与定理 2.2.7 又可看到，无变异的标准遗传算法总是收敛的，而且其收敛性在成熟度意义下还是单调的。然而，我们也注意到，作为标准遗传算法收敛的极限，可以是全局最优解，也可能是局部最优解，还可能是非极值解。这就揭示了杂交算子的二重性。它既可以强迫算法收敛，而又可能使算法过早终止到非最优解。这种现象称为遗传算法的早熟现象。

**定义 2.3.1** 称种群集 $A = \prod_{i=1}^{N} A_i \subset S^N$ 为早熟集，若对于任意 $\vec{X} \in A$ 和 $i \leqslant N$ 必有 $P\{T_c(T_s(\vec{X}))_i \in A_i\} = 1$。

显然 $S^N$ 是最大早熟集，$\prod_{i=1}^{N}\{X\}$ 为最小早熟集。

**定理 2.3.1** 若 $A = \prod_{i=1}^{N} A_i$ 为早熟集，必有 $A_i = A_j (i, j \leqslant N)$。

**证明** 对于任意 $X_1 \in A_1$，取

$$\vec{X} = (X_1, X_2, \cdots, X_N) \in A$$

则

$$P\{T_c(T_s(\vec{X}))_i = X_i\} \geqslant \frac{f^2(X_1)}{\left(\sum_{j=1}^{N} f(X_j)\right)^2} > 0$$

由定义 2.3.1 则证 $A_1 \subset A_i$。同样可证 $A_i \subset A_1$。则证 $A_1 = A_i (i \leqslant N)$。

**定理2.3.2**　若 $A \times A \times \cdots \times A$ 为早熟集，则 $A$ 必为包含 $A$ 的极小模式。

**证明**　设 $\mathscr{L}(A)$ 为包含 $A$ 的极小模式，则 $A \subset \mathscr{L}(A)$。下面证明 $\mathscr{L}(A) \subset A$。$\forall X \in \mathscr{L}(A)$，设 $A$ 的多样度为 $\lambda(A)$，成熟度为 $\beta(A)$，相应地极小模式为 $\mathscr{L}(a_{i_1}, \cdots, a_{i_{\lambda(A)}}; A)$，则 $\forall 1 \leqslant k \leqslant \lambda(A)$ 有 $x_{i_k} = a_{i_k}$，其中 $x_{i_k}$ 为 $X$ 的 $i_k$ 个基因分量。考虑分量 $I = \{1, 2l\} \setminus \{i_1, \cdots, i_{\lambda(A)}\}$，将 $I$ 重新编号，并记作 $\{j_1, j_2, \cdots, j_{\beta(A)}\}$。显然 $\forall 1 \leqslant k \leqslant \beta(A)$ 必存在 $Y_k \in A$ 使 $y_{kj_k} = x_{j_k}$，其中 $x_{j_k}$ 为 $X$ 的 $j_k$ 基因分量。取

$$\vec{Y}_1 = (Y_1, Y_2, \cdots, Y_2)$$

则

$$P\{T_c(T_s(\vec{Y}_1)) = Y_{12}\} > 0$$

其中 $Y_n$ 表示 $Y_1$ 与 $Y_2$ 在分量 $j_2$ 处杂交所得个体，于是 $Y_{12} \in A$。取

$$\vec{Y}_2 = (Y_{12}, Y_3, \cdots, Y_3)$$

则

$$P\{T_c(T_s(\vec{Y}_2)) = Y_{13}\} > 0$$

$Y_{13}$ 表示 $Y_{12}$ 在 $Y_3$ 在分量 $j_3$ 处杂交所得个体，从而 $Y_{13} \in A$，且 $Y_{13}$ 与 $X$ 的 $j_1, j_2, j_3$ 分量已经相同。递归下去直到 $\beta(A)$ 步，则 $Y_{1,\beta(A)} \in A$，且 $Y_{1,\beta(A)}$ 与 $X$ 的 $j_1, j_2, \cdots, j_{\beta(A)}$ 分量均相同，$Y_{1,\beta(A)} = X$，所以 $X \in A$，从而 $\mathscr{L}(A) \subset A$，结论得证。

遗传算法的早熟现象导致过早收敛。下面进一步研究导致过早收敛的原因。

**定义 2.3.2**　对于种群序列 $\{\vec{X}(n); n \geqslant 0\}$，若存在 $n$ 使

$d_j(\vec{X}(n)) = 0$，即 $x_{ij}(n) \equiv 1 (i \leqslant N)$ 或 $x_{ij}(n) \equiv 0$，称种群序列在 $j$ 基因上过早收敛。若 $d(\vec{X}(n)) = 0$，即 $d_j(\vec{X}(n)) = 0, (j \leqslant l)$，称种群序列过早收敛。

种群序列过早收敛等价于 $\vec{X}(n)$ 是齐次种群，因而等价于 $\beta(\vec{X}(n)) = 0$。

**定义 2.3.3** 对于种群 $\vec{X} \in S^N$，称

$$a_j(\vec{X}) = \frac{f_0(\vec{X})}{\bar{f}(\vec{X})}, \qquad b_j(\vec{X}) = \frac{f_1(\vec{X})}{\bar{f}(\vec{X})}, \tag{2.3.1}$$

分别为分量 $j$ 处 0 等位基因和 1 等位基因个体的适应比。

**定理 2.3.3** 设 $\{\vec{X}(n); n \geqslant 0\}$ 为标准遗传算法的种群马尔可夫链，$0 < p_m < 1$，则对于 $j \leqslant l$ 有

$$\begin{aligned} P\{d_j(\vec{X}(n+1)) = 0/\vec{X}(n) = \vec{X}\} \\ = (a_j(\vec{X}) + (1 - 2a_j(\vec{X}))p_m)^N \\ + (b_j(\vec{X}) + (1 - 2b_j(\vec{X}))p_m)^N \end{aligned} \tag{2.3.2}$$

**证明** 依定理 2.2.3 知

$$\begin{aligned} P\{x_{ij}(n+1) = 1/\vec{X}(n) = \vec{X}\} \\ = b_j(\vec{X}) + (1 - 2b_j(\vec{X}))p_m \\ P\{x_{ij}(n+1) = 0/\vec{X}(n) = \vec{X}\} \\ = a_j(\vec{X}) + (1 - 2a_j(\vec{X}))p_m \end{aligned}$$

由于 $(X_1(n+1), \cdots, X_N(n+1))$ 关于 $\vec{X}(n)$ 相互条件独立，于是

$$\begin{aligned} P\{d_j(\vec{X}(n+1)) = 0/\vec{X}(n) = \vec{X}\} \\ = P\{x_{ij}(n+1) = 0\ (i \leqslant N)/\vec{X}(n) = \vec{X}\} \\ + P\{x_{ij}(n+1) = 1\ (i \leqslant N)/\vec{X}(n) = \vec{X}\} \\ = [a_j(\vec{X}) + (1 - 2a_j(\vec{X}))p_m]^N \\ + [b_j(\vec{X}) + (1 - 2b_j(\vec{X}))p_m]^N \end{aligned}$$

则证。

**推论 2.3.1**　对于 $p_m = 0$ 的标准遗传算法的种群马尔可夫链 $\{\vec{X}(n); n \geqslant 0\}$ 有

$$P\{d_j(\vec{X}(n)) = 0 / \vec{X}(n) = \vec{X}\}$$
$$= a_j(\vec{X})^N + b_j(\vec{X})^N$$

**证明**　由定理 2.3.3 取 $p_m = 0$ 即得。

**推论 2.3.2**　设 $\{\vec{X}(n); n \geqslant 0\}$ 为标准遗传算法的种群马尔可夫链,记

$$g(\vec{X}, p_m, N)$$
$$= (a_j + (1 - 2a_j)p_m)^N + (b_j - (1 - 2b_j)p_m)^N$$
$$= (a_j + (1 - 2a_j)p_m)^N + (1 - a_j + (1 - 2a_j)p_m)^N$$

其中 $a_j = a_j(\vec{X}), b_j = b_j(\vec{X}) = 1 - a_j$。$g$ 为给定 $\vec{X}(n) = \vec{X}$ 时,$\vec{X}(n+1)$ 在分量 $j$ 处过早收敛的条件概率。则 $g$ 关于 $N$ 单调降,关于 $|p_m - 1/2|$ 和 $|a_j - 1/2|$ 单调增。特别当 $N$ 固定时,$g$ 在 $p_m = 1/2, a_j = 1/2$ 时达到最小值。

**证明**　由于

$$0 \leqslant a_j + (1 - 2a_j)p_m \leqslant 1$$

故 $g$ 关于 $N$ 单调降。令 $a = a_j + (1 - 2a_j)p_m$,则

$$\frac{\partial g}{\partial a_j} = Na^{N-1}(1 - 2p_m) - N(1 - a)^{N-1}(1 - 2p_m)$$
$$= N(1 - 2p_m)(a^{N-1} - (1 - a)^{N-1})$$

故

$$\frac{\partial g}{\partial a_j}\begin{cases} \leqslant 0, & 0 \leqslant a_j < 1/2 \\ \geqslant 0, & 1/2 \leqslant a_j \leqslant 1 \end{cases}$$

同理可得

$$\frac{\partial g}{\partial p_m}\begin{cases} \leqslant 0, & 0 \leqslant p_m < 1/2 \\ \geqslant 0, & 1/2 \leqslant p_m \leqslant 1 \end{cases}$$

则证。

**定义 2.3.4** 称 $F(\vec{X}) = \max\{f(X_i);\ i \leqslant N\}$ 为 $\vec{X} = (X_1, \cdots, X_N)$ 的种群满意值。称

$$M^* = \{\vec{X}; F(\vec{X}) = \max\{f(X);\ X \in S\}\}$$

为满意种群集。

改进的标准遗传算法有以下的步骤：

(1) 取 $k = 0$，选择初始种群 $\vec{X}(0)$；

(2) 利用标准遗传算法生成种群

$$\vec{X}'(k) = T_m \circ T_c \circ T_s(\vec{X}(k))$$

(3) 取适当的正常数 $\alpha$，计算

$$C_k = \min\left\{1; \left[\frac{\lambda(\vec{X}'(k))}{\lambda(\vec{X}(k))}\right]^{\alpha/k}\left[\frac{F(\vec{X}'(k))}{F(\vec{X}(k))}\right]^k\right\}$$

(4) 令 $\rho$ 是[0,1]上均匀分布的随机数，产生新种群

$$\vec{X}(k+1) = \begin{cases} \vec{X}'(k), & \rho \leqslant C_k \\ \vec{X}(k), & \rho > C_k \end{cases}$$

(5) 若符合停止准则停止，否则转向(2)。

**定理 2.3.4** 设 $\{\vec{X}(n); n \geqslant 0\}$ 是由改进的标准遗传算法产生的种群序列，则有：

(1) $\{\vec{X}(n); n \geqslant 0\}$ 是 $S^N$ 上非齐次的马尔可夫链；

(2) $\{\vec{X}(n); n \geqslant 0\}$ 以概率 1 收敛到满意种群集 $M^*$，即

$$\lim_{n\to\infty} P\{\vec{X}(n) \in M^* / \vec{X}(0)\} = 1$$

**证明** (1) 首先，$\{\vec{X}(k); k \geqslant 0\}$ 是马尔可夫链是显然的。状态转移矩阵为

$$P\{\vec{X}(k+1) = \vec{Y} / \vec{X}(k) = \vec{X}\}$$

$$= \begin{cases} P\{T_m \circ T_c \circ T_s(\vec{X}) = \vec{Y}\} \cdot C_k, & \vec{X} \neq \vec{Y} \\ 1 - \sum\limits_{\vec{Y} \neq \vec{X}} P\{\vec{X}(k+1) = \vec{Y} / \vec{X}(k) = \vec{X}\}, & \vec{X} = \vec{Y}. \end{cases}$$

因此是非齐次的马尔可夫链。

(2) 记

$$P_k\{\vec{X},\vec{Y}\} = P\{\vec{X}(k+1) = \vec{Y}/\vec{X}(k) = \vec{X}\}$$

$$P(k) = (P_k\{\vec{X},\vec{Y}\};\ \vec{X},\vec{Y} \in S^N)$$

由于

$$\lim_{k\to\infty} C_k = \begin{cases} 1, & F(\vec{Y}) \geqslant F(\vec{X}) \\ 0, & F(\vec{Y}) < F(\vec{X}) \end{cases}$$

则 $F(\vec{Y}) \geqslant F(\vec{X})$ 时有

$$\lim_{k\to\infty} P_k\{\vec{X},\vec{Y}\} = P\{T_m \circ T_c \circ T_s(\vec{X}) = \vec{Y}\} > 0$$

当 $F(\vec{Y}) < F(\vec{X})$ 时有

$$\lim_{k\to\infty} P_k\{\vec{X},\vec{Y}\} = 0$$

记

$$P(\infty) = \lim_{k\to\infty} P(k) = (P(\infty)(\vec{X},\vec{Y});\vec{X},\vec{Y} \in S^N)$$

则

$$P(\infty)(\vec{X},\vec{Y}) \begin{cases} > 0, & F(\vec{Y}) \geqslant F(\vec{X}) \\ = 0, & F(\vec{Y}) < F(\vec{X}) \end{cases}$$

显然 $P(\infty)$ 也是随机矩阵。于是 $P(\infty)$ 有唯一的一个不可约的非周期的正常返类 $M^*$。$S^N \setminus M^*$ 是非常返类。于是 $\{\vec{X}(k);k \geqslant 0\}$ 是强遍历的，对于任意初始状态 $\vec{X}(0) = \vec{X}_0$，有

$$\lim_{k\to\infty} P\{\vec{X}(k) = \vec{Y}/\vec{X}(0) = \vec{X}_0\} = \pi_\infty(\vec{Y})$$

且 $\sum_{\vec{Y}\in M} \pi_\infty(\vec{Y}) = 1$。于是

$$\lim_{k\to\infty} P\{\vec{X}(k) \in M^*/\vec{X}(0)\} = \sum_{\vec{Y}\in M^*} \pi_\infty(\vec{Y}) = 1$$

则证。

定理2.3.4指出，改进的标准遗传算法的关键是利用 $C_k$ 进行了控制。而 $C_k$ 是由多样度和种群满意适应值决定的。经过改进的标准遗传算法，或者种群满意值变大，或者种群多样度变大，都可

以使 $C_k$ 变大。也就是说,改进标准遗传算法的过程,或者增加种群满意值,或者增加种群搜索范围。当种群变化的过程增加时,增加搜索范围作用越来越小,而增加种群满意适应值越来越重要。因此在定理 2.3.4 中取代 $C_k$ 为

$$C_k = \min\left\{1; \left[\frac{F(\vec{X}'(k))}{F(\vec{X}(k))}\right]^k\right\}$$

同样成立。

# 2-4 优胜劣汰遗传算法的马氏链模型

杰出者遗传算法是另外一种改进的遗传算法。在标准遗传算法中取 $N$ 对母体进行杂交后变异产生新一代种群。而杰出者选择遗传算法是取 $(N-1)$ 对母体进行杂交后变异生成下一代种群的 $(N-1)$ 个个体,而最后一个个体取原种群中适应值最大的个体,这种修正保证了种群序列满意适应值的单调不减性。

假设 $S = \{0,1\}^l$ 为个体空间,$S^N$ 为种群空间,$S^2$ 为母体空间。用 $\{\vec{X}(n); n \geqslant 0\}$ 表示种群序列,$f: S \to R^+$ 为适应值函数,则杰出者选择遗传算法的过程为:

(1) 对于 $n = 0$ 给出初始种群 $\vec{X}(0)$;

(2) 在 $\vec{X}(n)$ 中独立的选择 $(N-1)$ 对母体 $(Y_1^{(k)}, Y_2^{(k)})$,其中 $k = 1 \sim (N-1)$;

(3) 对于(2)中选择的 $(N-1)$ 对母体 $(Y_1^{(k)}, Y_2^{(k)})$ $(k \leqslant N-1)$ 进行单点杂交得到 $(N-1)$ 个个体 $X_1'(n+1), \cdots, X_{N-1}'(n+1)$,通过变异产生 $(N-1)$ 个个体 $X_1(n+1), \cdots, X_{N-1}(n+1)$。

(4) 计算 $F(\vec{X}(n)) = \max\{f(X_j(n)); j \leqslant N\}$,若

$$f(X_{i_0}) = F(\vec{X}(n)) \tag{2.4.1}$$

即

$$i_0 = \arg\max_j \{f(X_j(n))\} \tag{2.4.2}$$

则 $X_N(n+1) = X_{i_0}(n)$。于是产生下一代种群：

$$\vec{X}(n+1) = (X_1(n+1), \cdots, X_{N-1}(n+1), X_N(n+1))^{\mathrm{T}}$$

(5) 若满足停机准则停止，否则转向(2)。

在杰出者选择遗传算法中使用的选择算子、杂交算子、变异算子与标准遗传算法相同。于是杰出者选择遗传算法可以记为

$$\vec{X}(n+1) = (X_i(n+1) = T_m^i \circ T_c^i \circ T_s^i(\vec{X}(n)), \quad (i \leqslant N-1); \quad X_N(n+1) = X_{i_0}(n)) \tag{2.4.3}$$

其中 $i_0$ 满足式(2.4.2)。

**定理 2.4.1**　杰出者选择遗传算法种群序列 $\{\vec{X}(n); n \geqslant 0\}$ 是有限齐次马尔可夫链。

**证明**　首先由于式(2.4.3)知 $\vec{X}(n+1)$ 仅与 $\vec{X}(n)$ 有关。因此 $\{\vec{X}(n); n \geqslant 0\}$ 是有限状态的马尔可夫链，且转移概率为

$$P\{\vec{X}(n+1) = \vec{Y}/\vec{X}(n) = \vec{X}\} = \begin{cases} \prod_{k=1}^{N} P\{T(\vec{X}(n))_k = Y_k\}, & \text{存在 } i_0 \in M(\vec{X}), \text{使 } Y_N = X_{i_0} \\ 0, & \text{否则} \end{cases} \tag{2.4.4}$$

其中

$$M(\vec{X}) = \{i;\ f(X_i) = \max_j \{f(X_j)\}\}$$

于是 $P\{\vec{X}(n+1) = \vec{Y}/\vec{X}(n)\}$ 与 $n$ 无关，因而是齐次的。转移概率可记为

$$P\{\vec{X}, \vec{Y}\} = P\{\vec{X}(n+1) = \vec{Y}/\vec{X}(n) = \vec{X}\} \tag{2.4.5}$$

**定理 2.4.2**　杰出者选择种群马尔可夫链序列 $\{\vec{X}(n); n \geqslant$

$0\}$ 以概率 1 收敛到满意种群集 $M^*$ 的子集 $M_0^*$：

$$M_0^* = \{\vec{Y} = (Y_1, \cdots, Y_N);\ Y_N \in M\}$$

即

$$\lim_{n\to\infty} P\{\vec{X}(n) \in M_0^* / \vec{X}(0) = \vec{X}_0\} = 1$$

**证明** 不妨假定 $X'$ 是 $f(X)$ 的唯一最大值解。由式(2.4.4)及式(2.4.5)知 $P\{\vec{X},\vec{Y}\}$ 有性质：

(1) 当 $\vec{X},\vec{Y} \in M_0^*$ 时，$P\{\vec{X},\vec{Y}\} > 0, P\{\vec{Y},\vec{X}\} > 0$，即 $\vec{X} \leftrightarrow \vec{Y}$

(2) 当 $\vec{X} \in M_0^*, \vec{Y} \notin M_0^*$ 时，$P\{\vec{X},\vec{Y}\} = 0$，即 $\vec{X} \nrightarrow \vec{Y}$。

于是 $M_0^*$ 为正常返的非周期的不可约闭集，$S^N \setminus M_0^*$ 为非常返的状态集。于是

$$\lim_{n\to\infty} P\{\vec{X}(n) = \vec{Y} / \vec{X}(0) = \vec{X}_0\} = \begin{cases} \pi(\vec{Y}), & \vec{Y} \in M_0^* \\ 0, & \vec{Y} \notin M_0^* \end{cases}$$

于是

$$\lim_{n\to\infty} P\{\vec{X}(n) \in M_0^* / \vec{X}(0) = \vec{X}_0\} = 1$$

**定理 2.4.3** 对于杰出者遗传算法马尔可夫链序列的种群满意值序列是单调不减的，即对于任意 $n \geqslant 0$ 有

$$F(\vec{X}(n)) \leqslant F(\vec{X}(n+1))$$

**证明** 由于杰出者遗传算法中取

$$X_N(n+1) = X_{i_0}(n)$$

其中 $i_0 = \arg\max_j \{f(X_j(n))\}$，则

$$F(\vec{X}(n)) \leqslant F(X_N(n+1)) \leqslant F(\vec{X}(n+1))$$

杰出者选择遗传算法与标准遗传算法，对于下一代种群前 $(N-1)$ 个个体取法相同，仅是最后一个个体取法不同。这样对于标准遗传算法的基本性质仍然成立。对于 $p_m = 0$ 时可得到下列定理。

**定理 2.4.4**　对于 $p_m = 0$ 的杰出者遗传算法有：

(1) 若 $Y \in \mathcal{A}(\dot{X}_0)$，必存在 $n > 0$ 使得

$$P\{Y \in \vec{X}(n) / \vec{X}(0) = \dot{X}_0\} > 0$$

(2) 对于 $Y \notin \mathcal{A}(\dot{X}_0)$ 和 $n \geqslant 0$ 有

$$P\{Y \in \dot{X}(n) / \dot{X}(0) = \dot{X}_0\} = 0$$

(3) 对于任意 $n \geqslant 1$ 有

$$P\{\vec{X}(n) \in H / \vec{X}(0) \in H\} = 1$$

(4) 多样度 $\{\lambda(\vec{X}(n)); n \geqslant 1\}$ 以概率 1 单调不增，即

$$P\{\lambda(\vec{X}(n+1)) \leqslant \lambda(\vec{X}(n))\} = 1$$

(5) 多样度 $\{\lambda(\vec{X}(n)); n \geqslant 1\}$ 以正概率单调减，即

$$P\{\lambda(\vec{X}(n+1)) < \lambda(\vec{X}(n))\} > 0$$

(6) 多样度 $\{\lambda(\vec{X}(n)); n \geqslant 1\}$ 以概率 1 收敛到 0，即

$$P\{\lim_{n \to \infty} \lambda(\vec{X}(n)) = 0\} = 1$$

**证明**　对于杰出者遗传算法，若已知 $\vec{X}(n)$，计算

$$i_0 = \arg\max_j \{f(X_j(n))\}$$

取选择算子为

$$P\{T_s^i(\vec{X}(n)\} = (X_{i_0}(n), X_{i_0}(n)) = 1 > 0$$

$T_s^i(i \leqslant N-1)$ 采用标准遗传算法的选择算子。于是原定理 2.2.3 和定理 2.2.4 证明均成立。

由此可见，杰出者选择遗传算法与标准遗传算法不同的是对一个体的产生采用了强制性选择。

与杰出者选择遗传算法相反的是稳定态遗传算法。它是用一对母体产生的个体去代替原种群适应值较小的个体。如果杰出者选择遗传算法采取了"优胜"的策略，那么稳定态遗传算法则采用了"劣汰"的策略。

稳定态遗传算法的基本步骤如下：

(1) 置 $n = 0$，随机产生初始种群 $\vec{X}(0) = (X_1(0), \cdots,$

$X_N(0)) \in S^N$。

(2) 从 $\vec{X}(n)$ 中选取一对母体 $(X_i(n), X_j(n))$，通过杂交和变异产生一个新个体 $X'(n+1)$。

(3) 按照以下概率

$$P\{T_d^\alpha(\vec{X}(n)) = \vec{X}(n) \setminus \{X_i(n)\}\} = \frac{f^{-\beta}(X_i(n))}{\sum_{k=1}^{N} f^{-\beta}(X_k(n))}$$

删去某一个体 $X_i(n)$ 用 $X'(n+1)$ 代之，得到新的种群

$$\vec{X}(n+1) = (X_1(n+1), \cdots, X_N(n+1))$$

(4) 若满足停机准则则停止，否则转向(2)。

**定理 2.4.5** 稳定态遗传算法是有限状态的齐次马尔可夫链，存在极限分布。

**证明** 记

$$P\{\vec{X}, \vec{Y}_i\} = P\{\vec{X}(n+1) = \vec{Y}_i / \vec{X}(n) = \vec{X}\}$$

其中

$$\vec{Y}_i = (X_1(n), \cdots, X_{i-1}(n), Y_i, X_{i+1}(n), \cdots, X_N(n))^T$$

于是

$$\begin{aligned} P\{\vec{X}, \vec{Y}_i\} = & \sum_{(Y_1, Y_2, Z) \in S^3} P\{T_s(\vec{X}(n)) = (Y_1, Y_2)\} \\ & \cdot P\{T_c(Y_1, Y_2) = Z\} \cdot P\{T_m(Z) = Y_i\} \\ & \cdot P\{T_d^\beta(\vec{X}(n) = \vec{X}(n) \setminus \{X_i\}\} \end{aligned} \tag{2.4.6}$$

显然 $P\{\vec{X}, \vec{Y}_i\} > 0$，且与 $n$ 无关，其它情况下 $P\{\vec{X}, \vec{Y}\} = 0$。于是稳定态遗传算法是有限状态齐次马尔可夫链。

对于任意 $\vec{Y} \in S^N$ 和 $\vec{X} \in S^N$ 有

$$P\{\vec{X}(N) = \vec{Y} / \vec{X}(0) = \vec{X}\} \geqslant \prod_{i=1}^{N} P\{\vec{X}, \vec{Y}_i\} > 0$$

于是 $P^N > 0$，于是转移概率 $P$ 为本原随机矩阵。根据定理2.1.7存在平稳分布：

$$\lim_{n\to\infty} P\{\vec{X}(n) = \vec{Y}/\vec{X}(0) = \vec{X}\} = \pi(\vec{Y})$$

由定理 2.2.4 可知，对于稳定态遗传算法可以无限次到达任何一个状态。因此，“劣汰”的策略不如“优胜”的策略。

**定理 2.4.6**　对于稳定态遗传算法采用变量删除因子，即

$$P\{T_d^{\beta_n}(\vec{X}(n)) = \vec{X}(n) - \{X_i\}\} = \frac{f^{-\beta_n}(X_i(n))}{\sum_{j=1}^{N} f^{-\beta_n}(X_j(n))}$$

则稳定态遗传算法是非时齐的马尔可夫链，且对于 $p_m > 0, \beta_n \to \infty$，$\{\vec{X}(n); n \geq 0\}$ 以概率 1 收敛到满意种群集，即

$$\lim_{n\to\infty} P\{\vec{X}(n) \in M^* / \vec{X}(0)\} = 1$$

**证明**　对于 $\vec{X}, \vec{Y} \in S^N$，有

$$\begin{aligned} P\{n, n+1, \vec{X}, \vec{Y}_i\} = & \sum_{(Y_1, Y_2, Z)} P\{T_s(\vec{X}(n)) = (Y_1, Y_2)\} \\ & \cdot P\{T_c(Y_1, Y_2) = Z\} \cdot P\{T_m(Z) = Y_i\} \\ & \cdot P\{T_d^{\beta_n}(\vec{X}(n)) = \vec{X}(n) \setminus \{X_i\}\} \end{aligned} \qquad (2.4.7)$$

其中 $\vec{Y}_i = (X_1, \cdots, X_{i-1}, Y_i, X_{i+1}, \cdots, X_N)$。当 $\vec{Y} \neq \vec{Y}_i (i \leq N)$ 时 $P\{n, n+1, \vec{X}, \vec{Y}\} = 0$。于是 $\{\vec{X}(n); n \geq 0\}$ 是非时齐的马尔可夫链。记

$$P_\infty(\vec{X}, \vec{Y}) = \lim_{n\to\infty} P\{n, n+1, \vec{X}, \vec{Y}\}$$

则有性质(2.4.7)及定理 1.2.6 知

$$P_\infty(\vec{X}, \vec{Y}) \begin{cases} > 0, & \vec{X} \in M^*, \vec{Y} \in M^* \\ > 0, & \vec{X} \notin M^*, \vec{Y} \in M^* \\ = 0, & \vec{X} \in M^*, \vec{Y} \notin M^* \end{cases}$$

且 $P_\infty(\vec{X}, \vec{X}) > 0$。因此

$$P(\infty) = (P_\infty(\vec{X}, \vec{Y}); \vec{X}, \vec{Y} \in S^N)$$

具有唯一的不可约非周期常返类 $M^*$。而 $S^N \setminus M^*$ 中的种群均为非常返状态。于是 $\{\vec{X}(n); n \geq 0\}$ 是强遍历的，即有

$$\lim_{n\to\infty} P\{\vec{X}(n) = \vec{Y}/\vec{X}(0)\} = \pi(\vec{Y}) \quad (\vec{Y} \in M^*)$$

且 $\sum_{\vec{Y}\in M^*} \pi(\vec{Y}) = 1$，于是

$$\begin{aligned}
&\lim_{n\to\infty} P\{\vec{X}(n) \in M^* / \vec{X}(0)\} \\
&= \lim_{n\to\infty} \sum_{\vec{Y}\in M^*} P\{\vec{X}(n) = \vec{Y}/\vec{X}(0)\} \\
&= \sum_{\vec{Y}\in M^*} \pi(\vec{Y}) = 1
\end{aligned}$$

则证.

我们知道，稳定态遗传算法是一种"劣汰"算法。但是删除算子的作用依赖于 $\alpha$。$\alpha$ 越小时"劣汰"是随机的，未必能达到"劣汰"的目的。当 $\alpha$ 越大时，"劣汰"才能真的实现。因此，只要"优胜"可以真的实现和"劣汰"可以真的实现，则 $\{\vec{X}(n); n \geqslant 0\}$ 必定收敛到满意解集。

如果在稳定态遗传算法中采取强迫淘汰制，一定淘汰最差的个体，即采用删除算子：

$$P\{T_d(\vec{X}(n)) = \vec{X}(n) \setminus \{X_i\}\} = \begin{cases} 1, & i = i_0 \\ 0, & i \neq i_0 \end{cases} \tag{2.4.8}$$

其中

$$i_0 = \min\{\arg\min_j \{f(X_j(n)\}\}$$

称为强迫"劣汰"遗传算法。

**定理 2.4.7** 强迫"劣汰"遗传算法是齐次马尔可夫链，且种群序列 $\{\vec{X}(n); n \geqslant 0\}$ 以概率 1 收敛到满意种群集 $M^*$，即

$$\lim_{n\to\infty} P\{\vec{X}(n) \in M^* / \vec{X}(0)\} = 1$$

**证明** 对于任意 $\vec{X}, \vec{Y} \in S^N$ 有

$$\begin{aligned}
P(n, n+1, \vec{X}, \vec{Y}_{i_0}) = &\sum_{(Y_1, Y_2, Z)} P\{T_s(\vec{X}(n)) = (Y_1, Y_2)\} \\
&\cdot P\{T_c(Y_1, Y_2) = Y_{i_0}\} \cdot P\{T_m(Z) = \vec{Y}_{i_0}\}
\end{aligned}$$

其中 $\vec{Y}_{i_0} = (X_1, \cdots, X_{i_0-1}, Y_{i_0}, X_{i_0+1}, \cdots, X_N)$，而

$$i_0 = \min\{\arg\min_{j \le N}\{f(X_j)\}\}$$

当 $\vec{Y} \neq \vec{Y}_{i_0}$ 时，$P(n, n+1, \vec{X}, \vec{Y}) = 0$。于是 $\{\vec{X}(n); n \geqslant 0\}$ 为齐次种群马尔可夫链。由于

$$P\{T_d(\vec{X}) = \vec{X} \setminus \{X_{i_0}\}\} = 1$$

于是 $j \neq i_0$ 时

$$P\{X_j \in T_d(\vec{X})\} = 1$$

类似定理 2.2.5 则证。

结合杰出者选择遗传算法和强制"劣汰"算法，我们可以得到"优胜劣汰"的遗传算法。它的计算步骤如下：

(1) 取 $n = 0$，随机选择初始种群 $\vec{X}(0)$；

(2) 在 $\vec{X}(n)$ 中独立地选择 $N$ 个母体，使用标准遗传算法得到种群

$$\vec{Y}(n) = (Y_1(n), \cdots, Y_N(n))$$

对于 $\vec{X}(n)$，记 $j_0 = \max\{\arg\min_j\{f(X_j(n))\}\}$。对于 $\vec{Y}(n)$ 记

$$i_0 = \min\{\arg\min_{j \le N}\{f(Y_j(n))\}\}$$

得到新一代种群

$$\vec{X}(n+1) = (Y_1, \cdots, Y_{i_0-1}, X_{j_0}(n), Y_{i_0+1}(n), \cdots, Y_N(n))$$

(3) 当满足停机准则时停止，否则转向(2)。

**定理 2.4.8**　"优胜劣汰"遗传算法是齐次马尔可夫链，且种群序列 $\{\vec{X}(n); n \geqslant 0\}$ 以概率 1 收敛到满意种群集 $M^*$，即

$$\lim_{n \to \infty} P\{\vec{X}(n) \in M^* / \vec{X}(0)\} = 1$$

**证明**　对于任意 $\vec{X}, \vec{Y} \in S^N$，当 $j \neq i_0$ 时

$$P_j(n, n+1, \vec{X}, \vec{Y}) = P\{X_j(n+1) = Y_j / \vec{X}\}$$

$$= \sum_{(Y_1^{(j)}, Y_2^{(j)}, Z_j) \in S} P\{T_s(\vec{X}) \doteq (Y_1^{(j)}, Y_2^{(j)})\}$$

$$P\{T_c(Y_1^{(j)},Y_2^{(j)})=Z_j\}\cdot P\{T_m(Z_j)=Y_j\}>0$$

而

$$P_{i_0}(n,n+1,\vec{X},\vec{Y})=P\{X_{i_0}(n+1)=Y_j/\vec{X}\}$$

$$=\begin{cases}1, & Y_j=X_{i_0}\\0, & Y_j\neq X_{i_0}\end{cases}$$

于是$\{\vec{X}(n);n\geqslant 0\}$是齐次种群马尔可夫链。由

$$P\{\vec{X},\vec{Y}\}=\prod_{j=1}^{N}P_j(n,n+1,\vec{X},\vec{Y})$$

利用定理2.3.1证明方法可证

$$\lim_{n\to\infty}P\{\vec{X}(n)\in M^*/\vec{X}(0)\}=1$$

# 2-5 等价类遗传算法的马氏链模型

设$S=\{0,1\}^l$为个体空间，$S^N$为种群空间。用函数$\varphi(\vec{X},X)$表示种群$\vec{X}$中个体$X$的个数，即

$$\varphi(\vec{X},X)=\sum_{i=1}^{N}\chi_{\{X\}}(X_i)$$

若对于任意$X\in S$，有$\varphi(\vec{X},X)=\varphi(\vec{Y},X)$，称$\vec{X}$和$\vec{Y}$等价，记作$\vec{X}\sim\vec{Y}$。

等价的种群$\vec{X}$和$\vec{Y}$有相同的结构，区别仅在于排列次序。这种排列次序实际上在遗传算法中是无特别作用的。因此我们可以将等价类作为一个种群来看。

如果假定$K=2^l$，$S$中元素记为$X_i(i\leqslant K)$记

$$\varphi_i=\varphi(\vec{X},X_i)\qquad(i\leqslant K)$$

$$(\varphi_1,\cdots,\varphi_k)=\{\vec{X};\varphi(\vec{X},X_i)=\varphi_i(i\leqslant K)\}$$

则将$S^N$分类得到新的等价类种群空间

$$S^{[K]}=\left\{(\varphi_1,\varphi_2,\cdots,\varphi_k);\sum_{i=1}^{K}\varphi_i=N\right\}$$

**定理 2.5.1**　对于等价类种群空间 $S^{[K]}$ 有

$$|S^{[K]}|<|S^N|$$

**证明**　由定理 1.1.2

$$|S^{[K]}|=\binom{N+K-1}{K-1}$$

且对于任意 $K>1,n>1$ 均有

$$nK>K+n-1$$

$$\prod_{n=1}^{N}(nK)>\prod_{n=1}^{N}(K+n-1)$$

$$N!K^N>K(K+1)\cdots(K+N-1)$$

于是

$$|S^N|=K^N>\frac{K(K+1)\cdots(K+N-1)}{N!}$$

$$=\binom{N+K-1}{K-1}=|S^{[K]}|$$

为方便起见,固定

$$S=\{X_i;\ i\leqslant K\}$$

$$S^{[K]}=\left\{(\varphi_1,\cdots,\varphi_k)=\varphi;\sum_{i=1}^{K}\varphi_i=N\right\}$$

其中

$$(\varphi_1,\cdots,\varphi_k)=\{\vec{X};\varphi(\vec{X},X_i)=\varphi_i(i\leqslant K)\}$$

$$P_{ijk}=P\{T_c(X_i,X_j)=X_k\}$$

$$P_{ij}'=P\{T_m(X_i)=X_j\}$$

有性质

$$P_{ijk}=P_{jik},\qquad P_{ij}'=P_{ji}',\qquad \sum_{k=1}^{N}P_{ijk}=1$$

对于 $\varphi \in S^{K!}$，$X_i \in S$，用 $P_{\varphi,i}$ 表示在等价类种群的条件下转移到第 $i$ 个个体 $X_i$ 的概率。于是有

$$
\begin{aligned}
& P\{\varphi(n+1) = (\varphi_1(n+1), \cdots, \varphi_K(n+1)) \\
& \quad / \varphi(n) = (\varphi_1(n), \cdots, \varphi_K(n))\} \\
& = \frac{N!}{\varphi_1(n+1)! \cdots \varphi_K(n+1)!} \prod_{i=1}^{K} (P_{(\varphi(n)),i})^{\varphi_i(n+1)}
\end{aligned}
\tag{2.5.1}
$$

在等价种群类 $\varphi(n)$ 的条件下，选择 $X_i$ 的概率为

$$
P\{T_s(\varphi(n))_i = X_i\} = \frac{\varphi_i(n) f(X_i)}{\sum_{j=1}^{K} \varphi_j(n) f(X_j)}
$$

记

$$
\varphi(n) \cdot f = \sum_{j=1}^{K} \varphi_j(n) f(X_j)
$$

则

$$
P\{T_s(\varphi(n))_i = X_i\} = \frac{\varphi_i(n) f(X_i)}{\varphi(n) f}
$$

于是

$$
\begin{aligned}
P_{\varphi(n),l} = & \sum_{i=1}^{K} \frac{\varphi_i(n) f(X_i)}{\varphi(n) f} \sum_{j=1}^{K} \frac{\varphi_j(n) f(X_j)}{\varphi(n) f} \\
& \cdot \sum_{r=1}^{K} P_{ijr} P_{rl}
\end{aligned}
\tag{2.5.2}
$$

如果对某一选择个体先变异后再杂交，有

$$
\begin{aligned}
P_{\varphi(n),l} = & \sum_{i=1}^{K} \frac{\varphi_i(n) f(X_i)}{\varphi(n) f} \sum_{j=1}^{K} \frac{\varphi_j(n) f(X_j)}{\varphi(n) f} \\
& \cdot \sum_{r=1}^{K} P_{ir} P_{rjl}
\end{aligned}
\tag{2.5.3}
$$

如果对两个选择的个体先变异后再杂交，则有

$$P_{\varphi(n),l} = \sum_{i=1}^{K} \frac{\varphi_i(n)f(X_i)}{\varphi(n)\bar{f}} \sum_{j=1}^{K} \frac{\varphi_j(n)f(X_j)}{\varphi(n)\bar{f}} \cdot \sum_{r=1}^{K}\sum_{s=1}^{K} P_{ir}P_{js}P_{rsl} \tag{2.5.4}$$

**定理 2.5.2** 对于 $0 < p_m < 1$，标准遗传算法的等价类种群序列是有限齐次不可约非周期的马尔可夫链。

**证明** 由于 $0 < p_m < 1$，则 $P_{ij} > 0 (i,j \leqslant K)$，且

$$P_{ji} > 0 \qquad (j \leqslant K, i \leqslant K)$$

取 $j \leqslant K$，使

$$P_{\varphi(n),j} = \frac{\varphi_j(n)f(X_j)}{\varphi(n)\bar{f}} > 0$$

由式(2.4.2) 即知

$$P_{\varphi(n),i} \geqslant P_{\varphi(n),j}P_{\varphi(n),j}P_{jjj}P_{ji} > 0$$

同理有式(2.4.3) 有

$$P_{\varphi(n),i} \geqslant P_{\varphi(n),j}P_{\varphi(n),j}P_{jji}P_{ii} > 0$$

由式(2.4.4) 也有

$$P_{\varphi(n),i} \geqslant P_{\varphi(n),j}P_{\varphi(n),j}P_{jjj}P_{jji}P_{ii} > 0$$

由式(2.4.1) 即得

$$P\{\varphi(n+1) = (\varphi_1(n+1),\cdots,\varphi_K(n+1)) / \varphi(n) = (\varphi_1(n),\cdots,\varphi_K(n))\} > 0 \tag{2.5.5}$$

由于 $\varphi(n+1)$ 仅与 $\varphi(n)$ 有关，再由式(2.5.1) 知式(2.5.5) 与 $n$ 无关，于是标准遗传算法的等价类种群序列是有限状态的齐次马尔可夫链。又由式(2.5.5) 知，$\{\varphi(n); n \geqslant 0\}$ 是不可约非周期的。

**定理 2.5.3** 对于标准遗传算法的等价类种群马尔可夫链 $\{\varphi(n); n \geqslant 0\}$，存在一种杂交概率 $P_{ijk} = 1/K (i,j,k \leqslant K)$，使得 $\{\varphi(n); n \geqslant 0\}$ 为平稳的马尔可夫链。

**证明** 由 $P_{ijk} = 1/K (i,j,k \leqslant K)$，即得 $P_{\varphi(n),i} = 1/K (i \leqslant K)$，于是

$$P\{\varphi(n+1)/\varphi(n)\} = \frac{N!}{\varphi_1(n+1)!\cdots\varphi_K(n+1)!}\frac{1}{K^N}$$

取

$$q(\varphi) = \frac{N!}{\varphi_1!\cdots\varphi_K!}\frac{1}{K^N}$$

则

$$\begin{aligned}
&\sum_{\varphi\in S^{(K)}} q(\varphi)P\{\varphi(n+1)/\varphi\} \\
&= \sum_{\varphi\in S^{(K)}}\left[\frac{N!}{\varphi_1!\cdots\varphi_K!}\frac{1}{K^N}\frac{N!}{\varphi_1(n+1)!\cdots\varphi_K(n+1)!}\frac{1}{K^N}\right] \\
&= \frac{N!}{\varphi_1!\cdots\varphi_K!}\frac{1}{K^N}P\{\varphi(n+1)\}
\end{aligned}$$

这时

$$\begin{aligned}
&P(\varphi_i(n+1)=j) \\
&= \sum_{\varphi:\varphi_i(n+1)=j}\frac{N!(N-j)!}{j!\prod\limits_{k\neq i}\varphi_k(n+1)!}\left(\frac{1}{K-1}\right)^{N-j} \\
&\quad\cdot\frac{(K-1)^{N-j}}{(N-j)!}\left(\frac{1}{K}\right)^j\left(\frac{1}{K}\right)^{N-j} \\
&= \frac{N!}{j!(N-j)!}\left(\frac{1}{K}\right)^j\left(\frac{K-1}{K}\right)^{N-j} \\
&= \binom{N}{j}\left(\frac{1}{K}\right)^j\left(1-\frac{1}{K}\right)^{N-j}
\end{aligned}$$

等价类种群标准遗传算法是遗传算法的缩写。它是把一个长度大小为 $N$ 的种群中的相同个体进行合并作为一个个体。$\varphi_i$ 不仅表示个体为 $X_i$，而且表示个体数目有 $\varphi_i$ 个。在标准遗传算法中，每个种群中的相同个体 $X_i$ 均以 $f(X_i)/(\varphi\cdot f)$ 的概率作出选择，而在等价类种群中个体 $X_i$ 统一用 $\varphi_i f(X_i)/(\varphi\cdot f)$ 的概率作出选择。杂交和变异与标准遗传算法相同。所不同的是在标准遗传算法中，种群相同的个体不是全部被杂交和变异。而在等价类种群中，

只要是相同个体一次性参加杂交与变异。因此等价类种群的标准遗传算法与一般标准遗传算法有相同的性质。例如等价类种群的标准遗传算法有极限分布，可以无限多次到达任何种群。

为了进一步研究等价类遗传算法的极限性质，需要具体研究等价类遗传算法的转移概率。

用 $S^{[k]}$ 表示等价类元素集合，转移概率为：

$$P\{s,s'\} = \frac{N!}{\prod_{X'\in S} s'(X')} \prod_{X'\in S} P\{s,X\}^{s'(x)} \tag{2.5.6}$$

其中 $s,s' \in S^{[k]}$，$s'(X)$ 表示 $S'$ 中含个体 $X$ 的数目，即 $\varphi(x)$。$P\{s,X\}$ 是由状态 $s$ 到个体 $X$ 的转移概率。若仅采用选择算子，则

$$P_1\{s,X\} = \frac{s(X)f(X)}{\sum_{Y\in S} s(Y)f(Y)} \tag{2.5.7}$$

同时采用选择算子与变异算子时有

$$P_2\{s,X\} = \sum_{y\in S} p_m^{d(X,Y)}(1-p_m)^{l-d(X,Y)} p_1(X,Y) \tag{2.5.8}$$

其中 $p_m$ 为变异概率。如果用 $P\{X\oplus Y,Z\}$ 表示通过 $X$ 与 $Y$ 杂交后变异为 $Z$ 的转移概率，则

$$P_3\{s,X\} = \sum_{X,Y\in S} P_1\{s,X\}P_2\{s,Y\}P\{X\otimes Y,Z\} \tag{2.5.9}$$

显然有

$$P_3\{X\otimes Y,Z\} = P\{(X\oplus Z)\otimes(Y\oplus Z),0\} \tag{2.5.10}$$

**定理2.5.4**　对于均匀杂交有

$$P\{X\otimes Y,0\} = \frac{1}{2}\sum_{k=0}^{|X\otimes Y|} p_c^k(1-p_c)^{|X\oplus Y|-k} \cdot \sum_{m=0}^{k}\binom{|(X\oplus Y)\otimes X|}{k-m}\binom{|(X\oplus Y)\otimes Y|}{m}$$

$$\cdot\left[p_m^{|X|+2m-k}(1-p_m)^{l-|X|-2m+k} + p_m^{|Y|-2m+k}(1-p_m)^{l-|Y|+2m-k}\right] \qquad (2.5.11)$$

**证明** 假定$k(0\leqslant k\leqslant|X\oplus Y|)$个基因位置利用均匀杂交交换，那么杂交后在$k$个位置上有不同值的概率为

$$p_c^k(1-p_c)^{|X\oplus Y|-k} \qquad (2.5.12)$$

在杂交之前，$|X\oplus Y|$个位置可以分为以下两种情况：

(1) 个体$X$与$Y$的基因分量为1和0的个数为$|(X\oplus Y)\otimes X|$；

(2) 个体$X$与$Y$的基因分量为0和1的个数为$|(X\oplus Y)\otimes Y|$。

不失一般性，假定$X$与$Y$的基因为1和0时有$(k-m)$个被杂交，而$X$与$Y$的基因为0和1时有$m$个被杂交，这一事件的概率为

$$\binom{|(X\oplus Y)\otimes X|}{k-m}\binom{|(X\oplus Y)\otimes Y|}{m}$$

通过均匀杂交，$X$的$|(X\oplus Y)\otimes X|$个位置上1的数目减少了$(k-m)$个，$Y$的$|(X\oplus Y)\otimes Y|$个位置上1的数目减少了$m$个。因此，$X$中1的数目增加了$(2m-k)$个，这样在$X'$中1的数目为$(|X|+2m-k)$，$X'$中的0数目为$(l-|X|-2m+k)$。类似地，在$Y'$中1的数目为$(|Y|-2m+k)$，$Y'$中0的数目为$(l-|Y|+2m-k)$。两个杂交个体中的一个变为0的概率为

$$\frac{1}{2}\left[p_m^{|X|+2m-k}(1-p_m)^{l-|X|-2m+k} + p_m^{|Y|-2m+k}(1-p_m)^{l-|Y|+2m-k}\right]$$

根据以上讨论的情况，对于$0\leqslant m\leqslant k$求和即得

$$\frac{1}{2}\sum_{m=0}^{k}\binom{|(X\oplus Y)\otimes X|}{k-m}\binom{|(X\oplus Y)\otimes Y|}{m}\cdot\left[p_m^{|X|+2m-k}(1-p_m)^{l-|X|-2m+k} + p_m^{|Y|-2m+k}(1-p_m)^{l-|Y|+2m-k}\right]$$

则证(2.5.11)。

**定理 2.5.5**　对于单点杂交有

$$
\begin{aligned}
P\{X \otimes Y, 0\} = & \frac{1-p_c}{2}\{p_m^{|X|}(1-p_m)^{l-|X|} \\
& + p_m^{|Y|}(1-p_m)^{l-|Y|}\} \\
& + \frac{p_c}{2}\sum_{k=1}^{l-1}\frac{1}{l-1}[p_m^{|X|+\Delta(X,Y,k)}(1-p_m)^{l-|X|-\Delta(X,Y,k)} \\
& + p_m^{|Y|+\Delta(X,Y,k)}(1-p_m)^{l-|Y|-\Delta(X,Y,k)}]
\end{aligned}
\tag{2.5.13}
$$

其中 $p_c$ 为杂交概率，且

$$
\Delta(X, Y, k) = |(2^k-1)| \odot |(2^k-1) \otimes Y|
$$

**证明**　类似定理 2.5.4。

**推论 2.5.1**　对于均匀杂交，$X, Y, Z \in S$，且满足

$$
|X \oplus Z| = 0, \qquad |Y \oplus Z| = l
$$

则有

$$
\begin{aligned}
P(X \otimes Y, Z) = \frac{1}{2}\{ & [p_c p_m + (1-p_c)]^l + [p_c(1-p_m) \\
& + (1+p_c)p_m]^l\}
\end{aligned}
\tag{2.5.14}
$$

**证明**　由假定知

$$
X \oplus Z = 0, \qquad Y \oplus Z = 2^l - 1
$$

于是有

$$
(X \oplus Z) \oplus (Y \oplus Z) = 2^l - 1
$$

于是

$$
[(X \oplus Z) \oplus (Y \oplus Z)] \otimes (X \otimes Z) = 0
$$

$$
[(X \oplus Z) \oplus (Y \oplus Z)] \otimes (Y \oplus Z) = 2^l - 1
$$

从而

$$
\begin{aligned}
P\{X \otimes Y, Z\} & = P\{(X \oplus Z) \times (Y \oplus Z), 0\} \\
& = \frac{1}{2}\sum_{k=0}^{l} p_c^k (1-p_c)^{l-k} \sum_{m=0}^{k}\binom{0}{k-m}\binom{l}{m}
\end{aligned}
$$

$$\cdot\left[p_m^{2m-k}(1-p_m)^{l-2m+k}+p_m^{l-2m+k}(1-p_m)^{2m-k}\right]$$

$$=\frac{1}{2}\sum_{k=0}^{l}p_c^k(1-p_c)^{l-k}\binom{l}{k}$$

$$\cdot\left[p_m^k(1-p_m)^{l-k}+p_m^{l-k}(1-p_m)^k\right]$$

$$=\frac{1}{2}\left[\sum_{k=0}^{l}\binom{l}{k}p_c^k(1-p_c)^{l-k}p_m^k(1-p_m)^{l-k}\right.$$

$$\left.+\sum_{k=0}^{l}\binom{l}{k}p_c^k(1-p_c)^{l-k}p_m^{l-k}(1-p_m)^k\right]$$

$$=\frac{1}{2}\{[p_cp_m+(1-p_c)(1-p_m)]^l$$

$$+[p_c(1-p_m)+(1-p_c)p_m]^l\}$$

**推论 2.5.2** 对于均匀杂交，$X,Y,Z\in S$，且满足

$$|X\oplus Z|=|Y\oplus Z|=l$$

则有

$$P\{X\otimes Y,Z\}=P\{X\oplus X,\bar{X}\}=p_m^l \tag{2.5.15}$$

**证明** 由假定知

$$X\oplus Z=2^l-1,\qquad Y\oplus Z=2^l-1$$

我们有

$$(X\oplus Z)\oplus(Y\oplus Z)=0$$

$$[(X\oplus Z)\oplus(Y\oplus Z)]\otimes(X\oplus Z)=0$$

$$[(X\oplus Z)\oplus(Y\oplus Z)]\otimes(Y\oplus Z)=0$$

于是

$$P\{X\otimes Y,Z\}=P\{((X\oplus Z)\otimes(Y\oplus Z),0\}$$

$$=\frac{1}{2}\sum_{k=0}^{l}p_c^k(1-p_c)^{-k}\sum_{m=0}^{k}\binom{0}{k-m}\binom{0}{m}$$

$$\cdot\left[p_m^{l+2m-k}(1-p_m)^{-2m+k}+p_m^{l-2m+k}(1-p_m)^{2m-k}\right]$$

$$=p_m^l$$

记

$$Q = (P(s, s'); s, s' \in S^{[k]})$$

$$q^{(n)} = (q^{n}[s]; s \in S^{[k]})$$

解线性方程组

$$q^{(n)} Q = q^{(n)}$$

得到

$$q^{(n)}[s] = \frac{\det(Q[s] - I)}{\sum_{s' \in S^{[k]}} \det(Q[s'] - I)} \tag{2.5.16}$$

其中 $Q[s]$ 是由 $Q$ 导出的，将 $Q$ 的 $s$ 行代替为 0 行向量。显然 $q^{(n)}$ 是 $Q$ 的平稳分布，且 $P_m$，$p_c$ 有关。记 $s[u]$ 为状态 $s' \in s^{[k]}$ 的集合，且满足条件：对某个 $Y \in S$，当 $u = u(X) \in U$ 时有：

(1) $s'(X) = N - 1$

(2) $s'(Y) = 1$

(3) $d(X, Y) = 1$

其中 $U \subset S^{[k]}$ 是一致种群 $u(X)(X \in S)$ 的集合。于是

$$q[s] = \lim_{p_m \to 0^+} q^{(m)}[s]$$

$$= \begin{cases} \dfrac{\lim\limits_{p_m \to 0^+} \det(Q^*[s] - I)}{\sum\limits_{s' \in U} \lim\limits_{p_m \to 0^+} \det(Q^*[s'] - I)}, & s \in U \\ 0, & s \in s^{[k]} \setminus U \end{cases} \tag{2.5.17}$$

其中 $Q^*[u](u \in U)$ 定义为$(s, s' \in S^k \setminus U + \{u\})$

$$Q^*[u](s, s')$$

$$= \begin{cases} 0, & s = u \\ Q(s, s') + \sum\limits_{u' \neq u} \delta(s' \in S[u']) \dfrac{Q(s, u')}{l}, & s \neq u \end{cases} \tag{2.5.18}$$

显然,对于任意 $s \in S^{(k)} \setminus U$,有

$$\sum_{s' \in S^{(k)} \setminus U \cup u} Q^*[u](s, s') = 1$$

**例 2.5.1** 设 $l = 2, N = 2$,则

$$S = \{(0\ 0)\ (0\ 1)\ (1\ 0)\ (1\ 1)\}$$

$$|S^{(k)}| = \binom{2 + 2^2 - 1}{2} = \binom{5}{2} = 10$$

于是 $S^{(k)}$ 中的元素为

| X | $S_0$ | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 0 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 1 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 0 |
| 1 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 2 | 1 | 0 |
| 1 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 2 |

从而得到

$$u(0\ 0) = s_0, \qquad u(0\ 1) = S_4$$

$$u(1\ 0) = s_7, \qquad u(1\ 1) = S_9$$

$$U = \{s_0, s_4, s_7, s_9\}$$

$$S[s_0] = \{s_1, s_2\}, \qquad S[s_4] = \{s_1, s_6\},$$

$$S[s_7] = \{s_2, s_8\}, \qquad S[s_9] = \{s_6, s_8\}$$

我们由式(2.5.18)可以计算出

$$Q^*[s_0](s_0\ s) = 0,\ s \in S^{(k)}$$

$$Q^*[s_0](s_1\ s_0) = P(s_1, s_0)$$

$$Q^*[s_1](s_1\ s_1) = P(s_1, s_1) + P(s_1, s_4)/2$$

**例 2.5.2** 取 $l = 1, N = 2, S = \{0,1\}, F = f(1)/f(0) < 1$,则

$$|S^{(k)}| = \binom{2 + 2 - 1}{1} = 3$$

记

$$[S^k] = \{s_0, s_1, s_2\}$$

| $i$ | $s_0$ | $s_1$ | $s_2$ |
|---|---|---|---|
| 0 | 2 | 1 | 0 |
| 1 | 0 | 1 | 2 |

转移概率矩阵为

$$Q = \begin{pmatrix} (1-p_m)^2 & 2p_m(1-p_m) & p_m^2 \\ \frac{((1-p_m)+Fp_m)^2}{(1+F)^2} & \frac{[(1-p_m)+Fp_m][p_m+F(1-p_m)]}{(1+F)^2} & \frac{[p_m+F(1-p_m)]^2}{(1+F)^2} \\ p_m^2 & 2p_m(1-p_m) & (1-p_m)^2 \end{pmatrix}$$

$$Q^*[s_0] = \begin{pmatrix} 0 & 0 \\ \frac{[(1-p_m)+Fp_m)^2}{(1+F)^2} & 1-\frac{[(1-p_m)+Fp_m]^2}{(1+F)^2} \end{pmatrix}$$

$$Q^*[s_2] = \begin{pmatrix} 1-\frac{[p_m+F(1-p_m)]^2}{(1+F)^2} & \frac{[p_m+F(1-p_m)]^2}{(1+F)^2} \\ 0 & 0 \end{pmatrix}$$

$$\det(Q^*[s_0]-I) = \frac{[(1-p_m)+Fp_m]^2}{(1+F)^2}$$

$$\det(Q^*[s_r]-I) = \frac{[p_m+F(1-p_m)]^2}{(1+F)^2}$$

由式(2.5.17)得到，当 $p_m \to 0$ 时有

$$q = (1/(1+F^2), 0, F^2/(1+F^2))$$

当 $F \to 0$ 时有

$$q = (1,0,0)$$

于是当 $p_m \to 0$ 及 $F \to 0$ 时，平稳分布集中在最优解种群上。

一般地，对于 $S$ 中个体适应值，假定

$$f(X_1)=f(X_2)=\cdots=f(X_k)>f(X_{k+1})\cdots>f(X_N)$$

$$F=\max_{f(X_i)\neq f(X_{i+1})}\frac{f(X_{i+1})}{f(X_i)}$$

当 $p_m\to 0$ 及 $F\to 0$ 时，平稳分布的极限分布集中在最优种群上。

## 2-6　遗传算法的马氏决策模型

设 $S$ 为个体空间，$S^N$ 为种群空间，$T_m$、$T_c$、$T_s$、$T_d$ 分别表示变异算子、杂交算子、选择算子、删除算子。我们考虑一般的稳定态遗传算法。

**定义 2.6.1**　称 $A\subset[0,1]\times[0,1]\times[0,\infty)\times[0,\infty)$ 为遗传算法的行为集。对于任意 $a=(a_1,a_2,a_3,a_4)\in A$，$a_1$ 代表变异概率，$a_2$ 代表杂交概率，$a_3$ 代表选择指数，$a_4$ 代表删除指数。

下面恒假定 $A$ 是有界闭集。

**定义 2.6.2**　称映射 $d_n: S^N\to A$ 为决策规则。若 $d_n$ 是确定性映射称为确定性决策规则；若 $d_n$ 为随机映射称为随机决策规则。

确定性规则的全体记为 $D^{md}$，随机性决策规则的全体记为 $D^{mr}$。

**定义 2.6.3**　称决策规则为遗传算法参数设置策略。若 $d_n=d(n\geqslant 0)$ 与 $n$ 无关，称 $\pi$ 是平稳的。若 $\pi$ 是平稳的且 $d$ 为常数，则称 $\pi$ 是简单的。

用 $\pi^{sd}$ 和 $\pi^{sr}$ 分别表示平稳的确定策略的全体和平稳的随机策略的全体。

在通常的遗传算法中采用的是简单的策略。

对于种群 $\vec{X}\in S^N$，记

$$F(\vec{X})=\max_{i\leqslant N}\{f(X_i)\}$$

给出决策参数 $a = (a_1, a_2, a_3, a_4) \in A$，得到

$$X = T_m^{a_1} \circ T_c^{a_2} \circ T_s^{a_3}(\vec{X}) \in S$$

$$Y = T_d^{a_4}(\vec{X}) \in S$$

$X$ 表示由 $\vec{X}$ 通过$(a_1, a_2, a_3)$策略，通过选择、杂交、变异而得到的一个个体。$Y$ 表示通过策略 $a_4$ 从 $\vec{X}$ 中删除的一个个体。

**定义 2.6.4**　对于二元函数

$$g(x, y) = \frac{x}{x + y} \tag{2.6.1}$$

称

$$r(\vec{X}, X, a) = g(F(\vec{X}), f(X)) \tag{2.6.2}$$

为瞬时损失函数，而称

$$\bar{r}(\vec{X}, a) = E(r(\vec{X}, X, a)) \tag{2.6.3}$$

为期望损失函数。若决策 $d: S^N \to A$ 是随机决策，有分布 $g_d$，则

$$\bar{r}_d(\vec{X}, a) = E(\bar{r}(\vec{X}, a)) \tag{2.6.4}$$

对于式(2.6.4)可以表示为

$$\bar{r}(\vec{X}, a) = \sum_{X \in S} g(F(\vec{X}), f(X)) p_a(\vec{X}, X) \tag{2.6.5}$$

其中

$$\begin{aligned} p_a(\vec{X}, X) &= P\{T_m^{a_1} \circ T_c^{a_2} \circ T_s^{a_3}(\vec{X}) = X\} \\ &= \sum_{Y \in S} \sum_{(X_1, X_2) \in S^2} P\{T_s^{a_3}(\vec{X}) = (X_1, X_2)\} \\ &\quad \cdot P\{T_c^{a_2}(X_1, X_2) = Y\} P\{T_m^{a_1}(Y) = X\} \end{aligned} \tag{2.6.6}$$

由式(2.6.5)及式(2.6.6)易见 $\bar{r}(\vec{X}, a)$ 是 $a$ 的连续函数。

**定义 2.6.5**　设 $P: S^N \times S^N \times A \to [0,1]$ 满足：对于任意 $\vec{X} \in S^N$ 及 $a \in A$，有

$$\sum_{\vec{Y} \in S^N} P\{\vec{X}, \vec{Y}, a\} = 1$$

称 $P$ 为转移概率矩阵族。

**定义 2.6.6** 称

$$\mathrm{MDP}_{GA} = \{S^N, S, A, r, P, \pi\}$$

为以$\pi$ 为策略的稳定态遗传算法决策模型。如果$\pi$ 是平稳的(或简单的),称 $\mathrm{MDP}_{GA}$ 为平稳的(或简单的)。

对于 $\mathrm{MDP}_{GA}$ 可以产生以下的遗传算法:

(1) 给出 $S^N$ 上的初始种群 $\vec{X}(0)$,置 $n = 0$;

(2) 按照 $d_n$ 的决策策略选择参数:

$$a(n) = (a_1(n), a_2(n), a_3(n), a_4(n))$$

(3) 根据参数 $a(n)$ 执行稳定态遗传算法:

$$X = T_m^{a_1(n)} \circ T_c^{a_2(n)} \circ T_s^{a_3(n)}(\vec{X})$$

$$Y = T_d^{a_4(n)}(\vec{X}(n))$$

(4) 在种群 $\vec{X}(n)$ 中以 $X$ 代替 $Y$;

(5) 计算瞬时损失 $r(\vec{X}(n), X, a(n))$;

(6) 若满足停机规则则停止;否则转向(2)。

**定理 2.6.1** $\mathrm{MDP}_{GA}$ 是马尔可夫链。

**证明** 由于 $\vec{X}(n+1)$ 仅依赖于 $a(n)$ 和 $\vec{X}(n)$ 的分布,而 $a(n)$ 也仅与 $\vec{X}(n)$ 的分布有关,因此$\{\vec{X}(n); n \geq 0\}$ 是马尔可夫链。

由于$\{\vec{X}(n); n \geq 0\}$ 为马尔可夫链,从而

$$y_n = r(\vec{X}(n), d_n(\vec{X}(n)))$$

$$d_n = d_n(\vec{X}(n))$$

是马尔可夫过程。

$\mathrm{MDP}_{GA}$ 一般是非齐次的马尔可夫链,若是简单的,则 $\mathrm{MDP}_{GA}$ 是齐次马尔可夫链。

**定理 2.6.2** 任给平稳的马尔可夫决策过程

$$\mathrm{MDP}_{GA} = \{S^N, S, A, r, P, \pi\}$$

以及 $S^N$ 上的初始分布 $P$,必存在概率空间$(\Omega, \beta(\Omega), P^\pi)$ 以及其

上以 $S^N$ 为状态空间的马尔可夫链 $\{\vec{X}(n);n\geqslant 0\}$，使得

$$P\{\vec{X}(n+1)=\vec{Y}/\vec{X}(n)=\vec{X}\}=P\{\vec{X},\vec{Y},d(\vec{X})\}$$

**证明**　令 $\Omega=(S^N)^*$，$\beta(\Omega)$ 为 $\Omega$ 的全体子集。用

$$\omega=(\omega_1,\omega_2,\cdots)\quad(\omega_i\in S^N)$$

表示 $\Omega$ 中的元素。对于任意 $k\geqslant 1$，$n_1<n_2<\cdots<n_k$，令

$$P^{\pi}_{n_1,\cdots,n_k}\{\omega_1,\omega_2,\cdots,\omega_{n_k}\}$$

$$=P\{\omega_{n_1}\}P\{\omega_{n_1},\omega_{n_2},d(\omega_{n_1})\}\cdots P\{\omega_{n_{k-1}},\omega_{n_k},d(\omega_{n_{k-1}})\}$$

则 $\{P^{\pi}_{n_1,\cdots,n_k}\}$ 为 $(\Omega,\beta(\Omega))$ 上相容的概率分布族。依照柯尔莫哥洛夫扩张定理知，该分布族在 $(\Omega,\beta(\Omega))$ 上唯一地确定了一个概率分布 $P^{\pi}$。对于任意 $n\geqslant 1$，令

$$\vec{X}_n(\omega)=\omega_n$$

易证 $\{\vec{X}(n);n\geqslant 0\}$ 是以 $P\{\vec{X},\vec{Y},d(\vec{X})\}$ 为转移概率的马尔可夫链。

**定义2.6.7**　给定有界实值函数 $V:S^N\to R^1$，$\vec{X}\in S^N$，$a\in A$，称

$$Q(\vec{X},a,V)=r(\vec{X},a)+\beta\sum_{\vec{Y}\in S^N}P\{\vec{X},\vec{Y},d(\vec{X})\}V(\vec{Y})\tag{2.6.7}$$

为 $Q$ 函数。其中 $0<\beta<1$。

**定理2.6.3**　对于 $S^N$ 上任意有界实值函数 $V$ 有

$$\min_{a\in A}Q(\vec{X},a,V)=\min_{g\in W}\sum_{a\in A}g(a)Q(\vec{X},a,V)\tag{2.6.8}$$

其中

$$W=\{g:A\to[0,1];g(a)\geqslant 0,\int_A g(a)da=1\}$$

**证明**　对于任意 $a\in A$，令

$$g(a')=\begin{cases}1, & a'=a\\0, & a'\neq a\end{cases}$$

则 $g \in W$,于是可证

$$\min_{a\in A} Q(\vec{X},a,V) \geqslant \min_{g\in W}\sum_{a\in A} g(a)Q(\vec{X},a,V)$$

另一方面对于任意 $g \in W$,有

$$\sum_{a\in A} g(a)Q(\vec{X},a,V) \geqslant \min_{a\in W} Q(\vec{X},a,V)$$

则证式(2.6.8)。

**定义 2.6.8** 设 $\beta \in (0,1)$ 为一个固定常数,称作折扣系数,以 $\pi = \{d_0, d_1, \cdots\}$ 为参数的 $MDP_{GA}$ 的值函数为

$$V_\pi(\vec{X}(0)) = \bar{r}(\vec{X}(0), d_0(\vec{X}(0))) + \sum_{n=1}^{\infty} \beta^n E(r(\vec{X}(0), d_n(\vec{X}(0))/\vec{X}(0))) \tag{2.6.9}$$

称 $V^*(\cdot) = \inf_\pi V_\pi(\cdot)$ 为 $MDP_{GA}$ 的最优值函数。称一个策略 $\pi^*$ 是最优的,如果有

$$V_{\pi^*}(\cdot) = V^*(\cdot)$$

由于定理 2.6.3,我们以后只须讨论 $\pi$ 为确定性策略,即 $\pi = \{d_n; n \geqslant 0\}$, $d_n \in MD$。因为 $0 \leqslant g(x,y) \leqslant 1$, $(x \geqslant 0, y \geqslant 0)$ 从而

$$|V_\pi(\cdot)| \leqslant \sum_{n=1}^{\infty} \beta^n = \frac{1}{1-\beta} < \infty$$

也即 $V_\pi(\cdot)$ 是 $S^N$ 上的有界实值函数。

进一步分析(2.6.9)式,取 $\pi = \{d_0, d_1, \cdots\}$ 为确定性决策,即 $d_n : S^N \to A$ 为确定性映射。则可记

$$r_{d_0}(\vec{X}(0)) = \bar{r}(\vec{X}(0), d_0(\vec{X}(0)))$$

是 $S^N$ 上的实数。而

$$E(r(\vec{X}(1), d_1(\vec{X}(1)))) = \sum_{X\in S} r_{d_1}(\vec{X}(1), X) P_{d_0}\{X/\vec{X}(1)\}$$

其中 $P_{d_0}\{X/\vec{X}(1)\}$ 表示从 $\vec{X}(0)$ 出发，由 $d_0(\vec{X}(0)) = a$ 策略产生 $\vec{X}(1)$，而在 $\vec{X}(1)$ 条件下由决策策略产生 $X$ 的条件概率。

于是有

$$E(r(\vec{X}(1), d_1(\vec{X}(1)))) = P_{d_0} r_{d_1}$$

其中 $P_{d_0}$ 为 $S \times S$ 上的矩阵，$r_{d_1}$ 为 $S$ 上的向量。一般地有

$$E(r(\vec{X}(n), d_n(\vec{X}(n)))) = P_{d_0} P_{d_1} \cdots P_{d_{n-1}} r_{d_n}$$

于是

$$V_\pi(\vec{X}(0)) = r_d(\vec{X}(0)) + \sum_{n=1}^{\infty} \beta^n P_{d_0} P_{d_1} \cdots P_{d_{n-1}} \cdot r_{d_n}(\vec{X}(0)) \tag{2.6.10}$$

对于平稳确定性决策有

$$V_\pi(\cdot) = r_d(\cdot) + \beta P_d V_\pi(\cdot) \tag{2.6.11}$$

**定理 2.6.4**　对于值函数 $V: S^N \to R^+$，当

$$V = L(V) = \min_{d \in MD}(r_d + \beta P_d V) \tag{2.6.12}$$

时有

$$V = V^* = \min_{d \in MD} V_\pi \quad (\pi = \{d\}) \tag{2.6.13}$$

**证明**　若 $V \leqslant L(V)$，则对于任意 $d \in MD$ 有

$$V \leqslant r_d + \beta P_d V$$

于是利用迭代方法即得

$$V \leqslant r_{d_0} + \beta P_{d_0} V \leqslant r_{d_0} + \beta P_{d_0}(r_{d_1} + \beta P_{d_1} V)$$
$$\leqslant r_{d_0} + \beta P_{d_0} r_{d_1} + \beta^2 P_{d_0} P_{d_1} V$$
$$\cdots$$

由于 $n$ 充分大时 $\sum_{h=n}^{\infty} \beta^h$ 可以充分小。于是对于任意 $\pi$ 有

$$V \leqslant V_\pi$$

从而 $V \leqslant V^* = \min_\pi V_\pi$。反之，若 $V \geqslant L(V)$，则有 $d'$ 使

$$V \geqslant r_d + \beta P_d V - \epsilon$$

对于任意 $\vec{X} \in S^N$ 成立，于是

$$(I - \beta P_d) V \geqslant (r_d - \epsilon e)$$

其中 $e$ 为 $S^N$ 上的向量，$I$ 为 $S^N \times S^N$ 上的单位矩阵。于是

$$\begin{aligned} V &\geqslant (I - \beta P_d)^{-1}(r_d - \epsilon e) \\ &\geqslant (I - \beta P_d)^{-1} r_d - (I - \beta_d)^{-1} \cdot \epsilon e \\ &\geqslant V_d - \epsilon' \end{aligned}$$

从而 $V \geqslant \min_{\pi} V_\pi = V^*$，于是满足式(2.6.12)的解必有 $V = V^*$。

**定理 2.6.5** 设 $V^*$ 是 $\mathrm{MPD}_{GA}$ 的最优值函数，则必存在一个平稳的确定性马氏策略 $\pi^* = (d^*)^\infty$ 使得

$$V_{\pi^*} = V^*$$

**证明** 由于 $\bar{r}(\vec{X}, a)$ 关于 $a$ 是连续函数，根据定理2.6.4知，$\arg\min_{a \in A} Q(\vec{X}, a, V^*) \neq \varnothing$。对于任意 $\vec{X} \in S^N$，令

$$d^*(\vec{X}) \in \arg\min_{a \in A} Q(\vec{X}, a, V^*)$$

则 $\pi^* = (d^*)^\infty$ 是一个平稳的确定性策略。由 $d^*$ 选取方法及定理 2.6.4 知

$$V^*(\vec{X}) = Q(\vec{X}, a, V^*)$$

下面证明 $V^* = V_{\pi^*}$。记

$$V^* = (V^*(\vec{X}); \vec{X} \in S^N)$$

则 $V^*$ 满足

$$\begin{aligned} V^*(\vec{X}) &= \bar{r}(\vec{X}, d^*(\vec{X})) \\ &\quad + \beta \sum_{\vec{Y} \in S^N} P(\vec{X}, \vec{Y}, d^*(\vec{X})) V^*(\vec{Y}) \end{aligned}$$

即

$$\begin{aligned} &(I - \beta P(\vec{X}, \vec{X}, d^*(\vec{X}))) V^*(\vec{X}) \\ &\quad - \beta \sum_{\vec{Y} \neq \vec{X}} P(\vec{X}, \vec{Y}, d^*(\vec{X})) V^*(\vec{Y}) \end{aligned}$$

$$= r(\vec{X}, d^*(\vec{X}))$$

将上式写成矩阵形式即有

$$(I - \beta\vec{P})V^* = r(d^*)$$

$$V^* = (I - \beta\vec{P})^{-1} r(d^*) = \sum_{n=0}^{\infty} \beta^n \vec{P}^n r(d^*)$$

由于 $\pi^* = (d^*)^{\infty}$，则对于任意 $\vec{X} \in S^N$ 有

$$V^*(\vec{X}) = r(\vec{X}, d^*) + \sum_{n=1}^{\infty} \beta^n (P_{d^*} \cdots P_{d^*} r_{d^*})(\vec{X}) = V_{\pi^*}(\vec{X})$$

则证。

**定理 2.6.6**　设 $\{Q_n(\vec{X}, a); n \geq 0\}$ 为下述迭代方法生成的 $Q$ 函数序列

$$Q_{n+1}(\vec{X}, a) = r(\vec{X}, a) + \beta \sum_{\vec{Y} \in S^N} [P\{\vec{X}, \vec{Y}, a\} \min_{a \in A} Q_n(\vec{X}, a)] \quad (2.6.14)$$

则对于任意初始 $Q_1(\vec{X}, a)$ 有

$$\lim_{n \to \infty} Q_n(\vec{X}, a) = Q^*(\vec{X}, a) \quad (\vec{X} \in S^N, a \in A)$$

**证明**　令

$$L(V(\vec{X})) = \min_{a \in A}\left(r(\vec{X}, a) + \beta \sum_{\vec{Y} \in S^N} P\{\vec{X}, \vec{Y}, a\} V(\vec{X})\right)$$

则由定理 2.6.4

$$L(V(\vec{X})) = V(\vec{X})$$

有最优解 $V^*(\vec{X})(\vec{X} \in S^N)$。于是对于

$$V_{n+1}(\vec{X}) = \min_{a \in A}\left\{r(\vec{X}, a) + \beta \sum_{\vec{Y} \in S^N} P\{\vec{X}, \vec{Y}, a\} V_n(\vec{Y})\right\}$$

有

$$\lim_{n \to \infty} V_{n+1}(\vec{X}) = V^*(\vec{X})$$

根据式(2.6.14)有

$$\lim_{n\to\infty} Q_n(\vec{X},a) = \bar{r}(\vec{X},a) + \beta\sum_{\vec{Y}\in S^N} P\{\vec{X},\vec{Y},a\} \lim_{n\to\infty} V_n(\vec{Y})$$
$$= \bar{r}(\vec{X},a) + \beta\sum_{\vec{Y}\in S^N} P\{\vec{X},\vec{Y},a\} V^*(\vec{Y})$$
$$= Q^*(\vec{X},a)$$

则证。

由定理 2.6.6 产生的迭代算法有两个实际困难：一是 $\bar{r}$ 与转移概率族必须已知，在 $MDP_{GA}$ 中，这些量无法用显式给出；二是每步迭代必须穷举 $S^N$ 中所有状态，这对 $MDP_{GA}$ 也是不现实的。为此下面给出 $MDP_{GA}$ 最优值函数的随机逼近方法。

即

$$Q_{n+1}(\vec{X},a) = Q_n(\vec{X},a) + \alpha(n)\chi_{(\vec{X},a)}(\vec{X}(n),a(n)) \cdot [Q_n(\vec{X}(n),a(n)) - \bar{r}(\vec{X},a) - \beta V_n(\vec{X}(n+1))] \tag{2.6.15}$$

其中 $\alpha(n) = 1/n^{\alpha}(1/2 < \alpha < 1)$，$a(n)$ 为 $A$ 上的均匀分布抽样，而

$$V_n(\vec{X}_{n+1}) = \min_{a\in A} Q_n(\vec{X}(n+1),a)$$

下面需要证明 $\{Q_n; n \geqslant 1\}$ 收敛到最优值函数 $Q^*$。

**定义 2.6.9** 考虑式(2.6.14) 中的马尔可夫链 $\{(\vec{X}(n), a(n)); n \leqslant 0\}$ 对于 $\vec{X} \in S^N$，$a \in A$，令 $n_0 = 0$。

$$n_k = \min\{n > n_{k-1};\ \vec{X}_n = \vec{X}, a(n) = a\}$$

称随机序列 $\{n_k; k \geqslant 1\}$ 为马尔可夫链对$(\vec{X},a)$ 的回访时间列。再令

$$\tau_k = n_{k+1} - n_k \quad (k \geqslant 1)$$

称随机序列 $\{\tau_k; k \geqslant 1\}$ 为马尔可夫链对$(\vec{X},a)$ 的回访时间的间隔列。

显然，式(2.6.14) 中的马尔可夫链 $\{(\vec{X}(n),a(n)); n \geqslant 0\}$

对任意状态的时间间隔列 $\{\tau_k; k \geqslant 1\}$ 是独立同分布的，且 $E(\tau_k) < \infty$。

**引理 2.6.1**　对于 $\alpha(n) = 1/n^{\alpha}(1/2 < \alpha < 1)$，任给正整数 $c > 0, d > 0$，令 $\bar{n}_k = [k^{1+\epsilon}](k \geqslant 1)$，其中 $[x]$ 表示小于等于 $x$ 的最大整数。定义子列

$$n_1 = d, \ n_k = n_{k-1} + c(\bar{n}_k - \bar{n}_{k-1}) \quad (k \geqslant 2)$$

则有

(1) $\sum_{n=1}^{\infty} \alpha(n) = \infty, \sum_{n \to \infty}^{\infty} \alpha^2(n) < \infty$

(2) $\lim_{n \to \infty}(n_k - n_{k-1}) = \infty$

(3) 当 $0 < \epsilon < \dfrac{1-\alpha}{\alpha}$ 时，有

$$\sum_{k=1}^{\infty} \alpha(n_k) = \infty, \sum_{k=1}^{\infty} \alpha^2(n_k) < \infty$$

**证明**　(1) 与(2) 显然。下证(3) 成立。首先，对于任意 $\epsilon > 0$，$\sum_{k=1}^{\infty} \alpha^2(n_k) < \infty$ 显然。依 $n_k$ 的定义

$$n_k = d + c\bar{n}_k - c\bar{n}_1 \qquad (k \geqslant 2)$$

因此

$$\sum_{k=1}^{\infty} \alpha(n_k) = \alpha(d) + \sum_{k=2}^{\infty} \alpha(d - c\bar{n}_1 + c\bar{n}_k)$$

由于 $0 < \epsilon < \dfrac{1-\alpha}{\alpha}$，则

$$(d - c\bar{n}_1 + c\bar{n}_k)^{\alpha} < (d - c\bar{n}_1 + ck^{1+\epsilon})^{\alpha}$$

$$\leqslant \left[c + (d - c\bar{n}_1)\frac{1}{k^{1+\epsilon}}\right]^{\alpha} k$$

于是

$$\sum_{k=2}^{\infty} \alpha(d - c\bar{n}_1 + c\bar{n}_k) = \infty$$

则证.

**引理 2.6.2** 取 $a(n) = 1/n^{\alpha}(1/2 < \alpha < 1)$，$A$ 为有限集，$\forall \vec{X} \in S^{N}$，$a \in A$，令

$$\alpha(n,\vec{X},a) = \alpha(n)\chi_{(\vec{X},a)}(\vec{X},a(n))$$

设$\{(\vec{X},a(n));n \geqslant 1\}$ 对$(\vec{X},a)$的回访时间间隔满足：对于任意 $M > 0$ 有

$$P\{\tau \leqslant M\} \geqslant \exp\left\{1 - \frac{1}{M^{l}}\right\}\left(l > \frac{\alpha}{1-\alpha}\right)$$

则依概率 1 有

$$\sum_{n=1}^{\infty}\alpha(n,\vec{X},a) = +\infty$$

$$\sum_{n=1}^{\infty}\alpha^{2}(n,\vec{X},a) < \infty$$

**证明** 设 $P\{n_k;k \geqslant 1\}$ 和$\{\tau_k;k \geqslant 1\}$ 分别为马尔可夫链对$(\vec{X},a)$的回访时间与回访时间间隔，则有

$$\sum_{n=1}^{\infty}\alpha(n_k) = \sum_{n=1}^{\infty}\alpha(n,\vec{X},a)$$

令

$$A = \left\{\omega;\sum_{n=1}^{\infty}\alpha(n_k) = +\infty\right\}$$

依引理 2.6.1 及 $\alpha(n)$ 的单调性，对任给的正整数 $d$，$c > 0$ 有

$$A \supset \{\omega;n_1 < d,\tau_k \leqslant c(\bar{n}_{k+1} - \bar{n}_k);k \geqslant 1\}$$

其中 $\bar{n}_k = [k^{1+\epsilon}]$。由于马尔可夫链$(\vec{X}(n),a(n))$是遍历的，故 $\forall\delta > 0$，取 $d$ 充分大使

$$P\{n_1 > d\} < \frac{1}{2}\delta$$

令 $Q_k = \bar{n}_{k+1} - \bar{n}_k(k \geqslant 1)$，由于 $l > \frac{\alpha}{1-\alpha}$，故 $A' = \sum_{k=1}^{\infty}\frac{1}{Q_k^{l}} < \infty$，从而 $\forall\delta > 0$，可取 $c$ 充分大使

$$\exp\left(-\frac{A'}{c^l}\right) > 1 - \frac{\delta}{2}$$

于是

$$P\{A\} \geqslant P\{\omega; n_1 \leqslant d, \tau_k \leqslant c(\bar{n}_{k+1} - \bar{n}_k); k \geqslant 1\}$$

$$\geqslant P\{\omega; \tau_k \leqslant c(\bar{n}_{k+1} - \bar{n}_k); k \geqslant 1\} - P\{n_1 > d\}$$

$$\geqslant \exp\left(-\frac{A'}{c^l}\right) - \frac{1}{2}\delta > 1 - \delta$$

由于 $\delta$ 的任意性即证 $P(A) = 1$。

**引理 2.6.3**　考虑随机迭代式

$$\Delta_{n+1}(x) = [1 - \alpha_n(x)]\Delta_n(x) + \alpha_n(x)F_n(x)$$

其中 $x \in A$，$A$ 为有限集。若满足

(1) $\forall x \in A, \sum_{n=1}^{\infty}\alpha_n(x) = \infty, \sum_{n=1}^{\infty}\alpha_n^2(x) < \infty$(a.e.)

(2) $\forall n \geqslant 1, \| E(F_n(x)) \| \leqslant \beta \| \Delta_n \| \quad (0 < \beta < 1)$

(3) $\forall n \geqslant 1, \| D(F_n(x)) \| \leqslant (C_1 + C_2 \| \Delta_n \| )^2 (C_1 > 0, C_2 > 0)$，则 $\forall x \in A, \lim_{n \to \infty}\Delta_n(x) = 0$(a.e.)。其中 $\| \Delta_n \| = \max_{x \in A}|\Delta_n(x)|$

**定理 2.6.7**　设式(2.6.14)中 $A$ 为有限集，马尔可夫链对 $(\vec{X}, a)$ 的回访时间间隔满足：

$$P\{\tau \leqslant M\} \geqslant \exp\left(-\frac{1}{M^l}\right) \left(l > \frac{\alpha}{1 - \alpha}\right)$$

则有

$$\lim_{n \to \infty} Q_n(\vec{X}, a) = Q(\vec{X}, a) \text{ (a.e.)}$$

**证明**　设 $Q^*(\vec{X}, a)$ 为 $\mathrm{MDP_{GA}}$ 最优 $Q$ 函数，$V$ 为最优值函数，于是

$$V^*(\vec{X}) = \min_{a \in A} Q^*(\vec{X}, a)$$

对于 $n \geqslant 1$，记

$$\Delta_n(x) = Q_n(\vec{X},a) - Q^*(\vec{X},a)$$
$$\alpha(n,\vec{X},a) = \alpha(n)\cdot\chi_{(\vec{X},a)}(\vec{X},a(n))$$
$$F_n(\vec{X},a) = [\vec{r}(\vec{X},a) + \beta V_n(\vec{X}(n+1) - Q^*(\vec{X},a)] \cdot\chi_{(\vec{X},a)}(\vec{X}(n),a(n))$$

则式(2.6.14) 可以写成

$$\Delta_{n+1}(\vec{X},a) = (1-\alpha(n,\vec{X},a))\Delta_n(\vec{X},a) - \alpha(n,\vec{X},a)F_n(\vec{X},a) \tag{2.6.16}$$

由引理 2.6.2 知,引理 2.6.3(1) 成立,下证(2) 和(3) 成立。

(2) $\forall n\geqslant 1,\vec{X}\in S^N,a\in A$,

$$|E(F_n(\vec{X},a))| = P\{\vec{X}(n)=\vec{X},a(n)=a\} \cdot\left[\vec{r}(\vec{X},a)+\beta\sum_{\vec{Y}\in S^N}P(\vec{X},\vec{Y},a)\cdot V_n(\vec{Y}) - Q^*(\vec{X},a)\right]$$
$$\leqslant\beta\sum_{\vec{Y}\in S^N}P(\vec{X},\vec{Y},a)|V_n(\vec{Y}) - V^*(\vec{Y})|$$
$$\leqslant\beta\sum_{\vec{Y}\in S^N}P(\vec{X},\vec{Y},a)\max_{a\in A}|Q(\vec{Y},a) - Q^*(\vec{Y},a)|$$
$$=\beta\|\Delta_n\|$$

(3) $\forall n\geqslant 1,\vec{X}\in S^N,a\in A$ 有

$$D(F_n(\vec{X},a))\leqslant E(F_n(\vec{X},a))$$
$$\leqslant 2E(r(\vec{X},a) - Q^*(\vec{Y},a))^2 + 2\beta^2\sum_{\vec{Y}\in S^N}P(\vec{X},\vec{Y},a)V_n^2(\vec{Y})$$

由于 $r(\vec{X},a)$ 为(0,1) 上的随机变量,于是

$$C = 2E(r(\vec{X},a) - Q^*(\vec{X},a))^2 < \infty$$

所以 $D(F_n(\vec{X},a))\leqslant C$

$$+ 2\beta^2\sum_{\vec{Y}\in S^N}P(\vec{X},\vec{Y},a)\cdot|V_n(\vec{Y}) - V^*(\vec{Y}) + V_n(\vec{Y})|^2$$
$$\leqslant C + 2\beta^2\sum_{\vec{Y}\in S^N}P(\vec{X},\vec{Y},a)[|V_n(\vec{Y}) - V^*(\vec{Y})|^2$$

$$+2|V_n(\vec{Y})-V^*(\vec{Y})|\cdot V^*(\vec{Y})+(V^*(\vec{Y}))^2]$$
$$\leqslant C+2\beta^2[\|\Delta_n\|^2+2d\|\Delta_n\|+d^2]$$

其中 $d=\max\limits_{\vec{Y}\in S^N}V^*(\vec{Y})$，则证引理2.6.3的(3)成立。于是本定理得证。

# 第 3 章

## 遗传算法收敛性的一般理论

### 3-1 遗传算法收敛的定义及性质

设 $E = S^N$ 为种群空间，$P$ 为$S^N$ 上的概率分布，$G_n : S^N \to S^N$ 为随机遗传算子，则

$$\vec{X}(n+1) = G_n(\vec{X}(n))$$

构成 $S^N$ 上的马尔可夫链。记

$$P^n\{\vec{X},\vec{Y}\} = P\{\vec{X}(n+1) = \vec{Y}/\vec{X}(n) = \vec{X}\}$$

则 $P^n\{\vec{X},\vec{Y}\}$ 为第 $n$ 步转移概率。对于任意 $B \subset S^N$，记

$$P^n\{\vec{X},B\} = \sum_{Y\in B} P^n\{\vec{X},\vec{Y}\}$$

称 $P^n\{\vec{X},B\}$ 为$\{\vec{X}(n)\}$ 的马尔可夫核。

设 $f: S \to R^+$ 为适应值函数，记全局最优解集为：

$$M = \{X;\ \forall\, Y \in S \text{ 有 } f(X) \geqslant f(Y)\}$$

**定义 3.1.1** 称$\{\vec{X}(n)\}$ 概率弱收敛到全局最优解集，若

$$\lim_{n\to\infty} P\{[\vec{X}(n) \cap M \neq \varnothing]\} = 1 \tag{3.1.1}$$

记作 $\vec{X}(n) \to M(\text{P.W.})$。

**定义 3.1.2**　称 $\{\vec{X}(n)\}$ 概率强收敛到全局最优解集，若

$$\lim_{n\to\infty} P\{[\vec{X}(n) \subset M]\} = 1 \tag{3.1.2}$$

记作 $\vec{X}(n) \to M(\text{P.S.})$。

**定义 3.1.3**　称 $\{\vec{X}(n)\}$ 几乎处处弱收敛到全局最优解集，若

$$P\{\lim_{n\to\infty}[\vec{X}(n) \cap M \neq \varnothing]\} = 1 \tag{3.1.3}$$

记作 $\vec{X}(n) \to M(\text{A.S.W})$。

**定义 3.1.4**　称 $\{\vec{X}(n)\}$ 几乎处处强收敛到全局最优解集，若

$$P\{\lim_{n\to\infty}[\vec{X}(n) \subset M]\} = 1 \tag{3.1.4}$$

记作 $\vec{X}(n) \to M(\text{A.S.S})$。

**定理 3.1.1**　种群序列 $\{\vec{X}(n)\}$ 的四种收敛性有关系：

(1) 几乎处处强收敛 $\Rightarrow$ 几乎处处弱收敛 $\Rightarrow$ 概率弱收敛；

(2) 几乎处处强收敛 $\Rightarrow$ 概率强收敛 $\Rightarrow$ 概率弱收敛。

**证明**　首先注意事实：

$$[\vec{X}(n) \subset M] \subset [\vec{X}(n) \cap M \neq \varnothing]$$

则

$$P\{[\vec{X}(n) \subset M]\} \leqslant P\{[\vec{X}(n) \cap M \neq \varnothing]\}$$

于是有式(3.1.2)$\Rightarrow$ 式(3.1.1)，式(3.1.4)$\Rightarrow$ 式(3.1.3)。另一方面，由于

$$\lim_{n\to\infty}[\vec{X}(n) \cap M \neq \varnothing] = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}[\vec{X}(n) \cap M \neq \varnothing]$$

于是由概率的单调性有

$$\begin{aligned}
&P\{\lim_{n\to\infty}[\vec{X}(n) \cap M \neq \varnothing]\} \\
&= \lim_{n\to\infty} P\{\bigcap_{k=n}^{\infty}[\vec{X}(n) \cap M \neq \varnothing]\} \\
&\leqslant \lim_{n\to\infty} P\{[\vec{X}(n) \cap M \neq \varnothing]\}
\end{aligned}$$

从而有式(3.1.3)$\Rightarrow$ 式(3.1.1)。同理有式(3.1.4)$\Rightarrow$ 式(3.1.2)。

**定义 3.1.5** 称 $B \subset S$ 为满意解集，若对于任意 $X \in B$，$Y \notin B$，均有 $f(X) > f(Y)$。

一个满意解集 $B$ 是这样的一个个体子集，它的每一个个体的适应值均大于满意解集之外的个体适应值。显然最优解集 $M$ 为满意解集。对于任意 $X \in S$，

$$B(X) = \{Y; f(Y) \geqslant f(X)\}$$

为满意解集。

若 $B_1, B_2$ 为满意解集，则仅有以下两种情况，即 $B_1 \subset B_2$ 或 $B_2 \subset B_1$，于是 $B_1 \cup B_2$，$B_1 \cap B_2$ 均为满意解集。全局最优解集 $M$ 即是所有满意解集之交集，$M$ 为最小满意解集。

定义 3.1.1 ~ 3.1.4 中可以用满意解集 $B$ 代替最小解集 $M$，这时称为在对应意义下收敛到满意解集。特别是对于 $\varepsilon > 0$ 有满意解集：

$$B(\varepsilon) = \{Y; f(Y) \geqslant f(X^*) - \varepsilon\}$$

其中 $X^*$ 表示最优解，即 $X^* \in M$。在实际问题中只要求收敛到满意解集 $B(\varepsilon)$ 即可。在理论上来讲，如果 $\{\vec{X}(n)\}$ 对于任意满意解集是收敛的，则必收敛到全局最优解集。

对于 $\vec{X} = (X_1, \cdots, X_N) \in S^N$，记

$$B(\vec{X}) = \{X_i; f(X_i) \geqslant f(X_j)\ (j \leqslant N)\} \subset S$$

$$D(\vec{X}) = \{X_i; f(X_i) \leqslant f(X_j)\ (j \leqslant N)\} \subset S$$

若 $\{\vec{X}(n)\}$ 是遗传算法序列，它是 $S^N$ 上的马尔可夫链，取

$$b_n = b(\vec{X}(n)) \in B(\vec{X}(n))$$

$$d_n = d(\vec{X}(n)) \in D(\vec{X}(n))$$

则 $\{b_n\}$ 及 $\{d_n\}$ 均为 $S$ 上的马尔可夫链；称 $\{b_n\}$ 为 $\{\vec{X}(n)\}$ 的优值马尔可夫链，$\{d_n\}$ 为 $\{\vec{X}(n)\}$ 的劣值马尔可夫链。

**定义 3.1.6** 称 $\{\vec{X}(n)\}$ 的优值马尔可夫链 $\{b_n\}$ 概率收敛到全局最优解集，若

$$\lim_{n \to \infty} P\{b_n \in M\} = 1$$

称 $\{b_n\}$ 几乎处处收敛到全局最优解集，若

$$P\{\lim_{n\to\infty}[b_n \in M]\} = 1$$

**定义 3.1.7**　称 $\{\vec{X}(n)\}$ 的劣值马尔可夫链 $\{d_n\}$ 概率收敛到全局最优解集，若

$$\lim_{n\to\infty} P\{d_n \in M\} = 1$$

称 $\{d_n\}$ 几乎处处收敛到全局最优解集，若

$$P\{\lim_{n\to\infty}[d_n \in M]\} = 1$$

**定理 3.1.2**　对于 $\{\vec{X}(n)\}$ 及其优值马尔可夫链 $\{b_n\}$ 和劣值马尔可夫链 $\{d_n\}$ 有以下关系：

(1) $\{\vec{X}(n)\}$ 概率弱收敛到全局最优解集，当且仅当存在优值马尔可夫链 $\{b_n\}$ 概率收敛到全局最优解集。

(2) $\{\vec{X}(n)\}$ 几乎处处弱收敛到全局最优解集，当且仅当存在优值马尔可夫链 $\{b_n\}$ 几乎处处收敛到全局最优解集。

(3) $\{\vec{X}(n)\}$ 概率强收敛到全局最优解集，当且仅当存在劣值马尔可夫链 $\{d_n\}$ 概率收敛到全局最优解集。

(4) $\{\vec{X}(n)\}$ 几乎处处强收敛到全局最优解集，当且仅当存在劣值马尔可夫链 $\{d_n\}$ 几乎处处收敛到全局最优解集。

**证明**　由于 $b_n \in M$ 必有 $\vec{X}(n) \cap M \neq \varnothing$，而 $\vec{X}(n) \cap M \neq \varnothing$ 时 $B(\vec{X}(n)) \cap M \neq \varnothing$，从而存在 $b_n \in B(\vec{X}(n)) \cap M, b_n \in M$。则证(1)与(2)。又因为 $d_n \in M$ 必有 $\vec{X}(n) \subset M$，而 $\vec{X}(n) \subset M$ 时，$D(\vec{X}(n)) \subset M$，从而存在 $d_n \in D(\vec{X}(n)) \cap M, d_n \in M$。则证(3)与(4)。

我们指出定理 3.1.2 对于满意解集 $B$ 同样成立。这时 $\{b_n\}$ 与 $\{d_n\}$ 相应地收敛到满意解集 $B$。

$\{\vec{X}(n)\}$ 是 $S^N$ 上的马尔可夫链时，$\{b_n\}$ 和 $\{d_n\}$ 是 $S$ 上的马尔可夫链，则 $f(b_n)$ 与 $f(d_n)$ 是 $S$ 上的实值马尔可夫链。这样用

$\{b_n\}$ 与 $\{d_n\}$，或 $\{f(b_n)\}$ 与 $\{f(d_n)\}$ 讨论 $\{\vec{X}(n)\}$ 的收敛性会更方便。

**定义 3.1.8** 设 $f_n$ 是 $S$ 上的实值随机变量序列，称 $f_n$ 概率收敛到 $f^*$，若对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\{|f_n - f^*| \leqslant \varepsilon\} = 1$$

记作 $f_n \to f^*(P)$。称 $f_n$ 几乎处处收敛到 $f^*$，若

$$P\{\lim_{n\to\infty} f_n = f^*\} = 1$$

记作 $f_n \to f^*(\text{a.s.})$。

**定理 3.1.3** $\{\vec{X}(n)\}$ 概率弱收敛到最优解集，当且仅当存在优值马尔可夫链 $\{b_n\}$，它的适应值序列 $\{f(b_n)\}$ 概率收敛到最优适应值 $f^*$，即

$$\vec{X}(n) \to M(\text{P.W.}) \Leftrightarrow \exists\, b_n \in B(\vec{X}(n)), f(b_n) \to f^*(P)$$

其中

$$f^* = \max\{f(X);\ x \in S\}。$$

**证明** 由于 $S$ 是有限的，在 $S$ 上的适应值全体 $f$ 也是有限的。于是当 $n$ 充分大时有

$$\begin{aligned}\vec{X}(n) \cap M \neq \varnothing &\Leftrightarrow \exists\, b_n \in B(\vec{X}(n)), b_n \in M\\ &\Leftrightarrow \exists\, b_n \in B(\vec{X}(n)), f(b_n) = f^*\end{aligned}$$

则证。

**定理 3.1.4** $\{\vec{X}(n)\}$ 概率强收敛到最优解集，当且仅当存在劣值马尔可夫链 $\{d_n\}$，它的适应值序列 $\{f(d_n)\}$ 概率收敛到最优适应值 $f^*$，即

$$\begin{aligned}&\vec{X}(n) \to M(\text{P.S.})\\ &\Leftrightarrow \exists\, d_n \in D(\vec{X}(n)), f(d_n) \to f^*(P)\end{aligned}$$

**证明** 类似于定理 3.1.3。

**定理 3.1.5** $\{\vec{X}(n)\}$ 几乎处处弱收敛到最优解集，当且仅当存在优值马尔可夫链 $\{b_n\}$，其适应值序列 $\{f(b_n)\}$ 几乎处处收敛

到最优适应值 $f^*$，即

$$\vec{X}(n) \to M(\mathrm{A.S.W})$$
$$\Leftrightarrow \exists\, b_n \in B(\vec{X}(n)), f(b_n) \to f^*\ (\mathrm{a.s.})$$

$\{\vec{X}(n)\}$ 几乎处处强收敛到最优解集，当且仅当存在劣值马尔可夫链 $\{d_n\}$，其适应值序列 $\{f(d_n)\}$ 几乎处处收敛到最大适应值 $f^*$，即

$$\vec{X}(n) \to M(\mathrm{A.S.S})$$
$$\Leftrightarrow \exists\, d_n \in d(\vec{X}(n)), f(d_n) \to f^*\ (\mathrm{a.s.})$$

**证明**　类似于定理 3.1.3。

对于种群 $\vec{X} = (X_1, \cdots, X_N)$，称

$$\bar{f}(\vec{X}) = \frac{1}{N}\sum_{i=1}^{N} f(X_i)$$

为种群适应值的平均值。对于马尔可夫链 $\{\vec{X}(n)\}$

$$\bar{f}_n = \bar{f}(\vec{X}(n)) = \frac{1}{N}\sum_{i=1}^{N} f(X_i(n))$$

为种群平均适应值序列，称为种群均值马尔可夫链。

**定理 3.1.6**　$\{\vec{X}(n)\}$ 几乎处处强收敛到最优解集当且仅当它相应的种群均值马尔可夫链几乎处处收敛到最优值，即

$$\vec{X}(n) \to M(\mathrm{A.S.S}) \Leftrightarrow \bar{f}(\vec{X}(n)) \to f^*\ (\mathrm{a.s.})$$

$\{\vec{X}(n)\}$ 概率强收敛到最优解集当且仅当它相应的种群均值马尔可夫链概率收敛到最优值，即

$$\vec{X}(n) \to M(\mathrm{P.S}) \Leftrightarrow \bar{f}(\vec{X}(n)) \to f^*\ (P)$$

**证明**　由于 $S^N$ 中元素有限，则 $\{\bar{f}_n; \bar{f}_n = \bar{f}(\vec{X}(n))\}$ 为有限集。于是 $\bar{f}_n = f^*$ 等价于 $\vec{X}(n) \subset M$（当 $n$ 充分大时），则证。

## 3-2　遗传算法概率收敛定理

设 $P$ 为 $S^N$ 上的概率分布，$\{\vec{X}(n)\}$ 为 $S^N$ 上的种群马尔可夫

链序列，$M$ 为全局最优解集，$B$ 为任意满意解集，记

$$\alpha_n^B = P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) \cap B \neq \varnothing\} \tag{3.2.1}$$

$$\beta_n^B = P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) \cap B = \varnothing\} \tag{3.2.2}$$

则有以下定理。

**定理 3.2.1** 若 $\{\alpha_n^B\}$，$\{\beta_n^B\}$ 满足：

(1) $\sum_{n=1}^{\infty}(1-\beta_n^B) = \infty$

(2) $\lim_{n\to\infty} \frac{\alpha_n^B}{1-\beta_n^B} = 0$

则 $\{\vec{X}(n)\}$ 概率弱收敛到满意解集 $B$，即

$$\lim_{n\to\infty} P\{\vec{X}(n) \cap B \neq \varnothing\} = 1$$

若对任意满意解集 $B$ 概率弱收敛，则 $\{\vec{X}(n)\}$ 概率弱收敛到全局最优解集 $M$。

**证明** 记

$$P_0(n) = P\{\vec{X}(n) \cap B = \varnothing\}$$

则由贝叶斯公式有

$$\begin{aligned}
P_0(n+1) &= P\{\vec{X}(n+1) \cap B = \varnothing\} \\
&= P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) \cap B \neq \varnothing\} \\
&\quad \cdot \{P\{\vec{X}(n) \cap B \neq \varnothing\} \\
&\quad + P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) \cap B = \varnothing\} \\
&\quad \cdot \{P\{\vec{X}(n) \cap B = \varnothing\} \\
&\leqslant \alpha_n^B + \beta_n^B P_0(n)
\end{aligned} \tag{3.2.3}$$

由条件(2)，对于任意 $\varepsilon > 0$，存在 $N_1$，当 $n \geqslant N_1$ 时有

$$\frac{\alpha_n^B}{1-\beta_n^B} \leqslant \frac{\varepsilon}{2}$$

即 $\alpha_n^B \leqslant \frac{\varepsilon}{2}(1-\beta_n^B)$，由式(3.2.3)即得

$$\left(P_0(n+1)-\frac{\varepsilon}{2}\right)-\beta_n^B\left(P_0(n)-\frac{\varepsilon}{2}\right)$$
$$\leqslant P_0(n+1)-\beta_n^B P_0(n)-\alpha_n^B \leqslant 0$$

从而

$$P_0(n+1)-\frac{\varepsilon}{2} \leqslant \beta_n^B\left(P_0(n)-\frac{\varepsilon}{2}\right) \tag{3.2.4}$$

利用递归公式(3.2.4)即得

$$P_0(n+1) \leqslant \frac{\varepsilon}{2}+\prod_{k=1}^{n}\beta_n^B\left(P_0(n)-\frac{\varepsilon}{2}\right)$$

条件(1)等价于 $\prod_{k=1}^{\infty}\beta_n^B=0$，于是存在 $N_2$，当 $n \geqslant N_2$ 时有 $\prod_{k=1}^{n}\beta_k^B \leqslant \frac{\varepsilon}{2}$，从而 $n \geqslant \max(N_1,N_2)$ 时有 $P_0(n+1) \leqslant \varepsilon$，由 $\varepsilon$ 的任意性即得到：

$$\lim_{n\to\infty}P\{\vec{X}(n)\cap B=\varnothing\}=\lim_{n\to\infty}P_0(n)=0$$

于是有

$$\lim_{n\to\infty}P\{\vec{X}(n)\cap B\neq\varnothing\}=1$$

由定理3.2.1可知，$\vec{X}(n)$ 的收敛性与 $\alpha_n^B$ 与 $\beta_n^B$ 参数关系密切。$\alpha_n^B$ 是指在 $\vec{X}(n)$ 已进入满意解集时下一步脱离满意解集的概率，显然 $\alpha_n^B$ 应是越来越小。$\beta_n^B$ 是 $\vec{X}(n)$ 未进入满意解集时下一步仍不能进入满意解集的概率，这个概率就不能太小。特别当 $\beta_n^B=\beta<1$ 是常数，$\lim\limits_{n\to\infty}\alpha_n^B=0$ 时满足定理条件。显然 $\beta_n^B$ 可以为0，但是 $\beta_n^B\neq 1$，这在直观上也是对的。如果 $\beta_n^B=1$，即是说 $\vec{X}(n)$ 未进入满意解集，则下一步必定不进入满意解集，这样 $\{\vec{X}(n)\}$ 不可能收敛。

为了给出概率强收敛定理，记

$$\bar{\alpha}_n^B=P\{\vec{X}(n+1)\cap B^c\neq\varnothing/\vec{X}(n)\cap B^c=\varnothing\}$$

$$\beta_n^B = P\{\vec{X}(n+1) \cap B^c \neq \varnothing / \vec{X}(n) \cap B^c \neq \varnothing\}$$

**定理 3.2.2** 若 $\{\bar{\alpha}_n^B\}$，$\{\bar{\beta}_n^B\}$ 满足

(1) $\sum\limits_{n=1}^{\infty}(1-\bar{\beta}_n^B) = \infty$

(2) $\lim\limits_{n\to\infty} \dfrac{\bar{\alpha}_n^B}{1-\bar{\beta}_n^B} = 0$

则 $\{\vec{X}(n)\}$ 概率强收敛到满意解集 $B$，即

$$\lim_{n\to\infty} P\{\vec{X}(n) \subset B\} = 1$$

若对任意满意解集 $B$，$\{\vec{X}(n)\}$ 概率强收敛，则 $\{\vec{X}(n)\}$ 概率强收敛到全局最优解集 $M$。

**证明** 记

$$P_0(n) = P\{\vec{X}(n) \cap B^c \neq \varnothing\}$$

则由贝叶斯公式有

$$P_0(n+1) \leqslant \bar{\alpha}_n^B + \bar{\beta}_n^B P_0(n)$$

仿定理 3.2.1 证明过程可证。

**定理 3.2.3** 若对于任意满意解集 $B$ 有：

$$P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) = \vec{X}\} \leqslant a_n \quad (\vec{X} \cap B \neq \varnothing)$$

$$P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) = \vec{X}\} \leqslant b_n \quad (\vec{X} \cap B = \varnothing)$$

且满足条件：

(1) $\sum\limits_{n=1}^{\infty}(1-b_n) = \infty$

(2) $\dfrac{a_n}{1-b_n} \to 0$

则 $\{\vec{X}(n)\}$ 概率弱收敛到全局最优解集 $M$。

**证明** 由于

$$\alpha_n^B = P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) \cap B \neq \varnothing\}$$
$$= \frac{P\{\vec{X}(n+1) \cap B = \varnothing, \vec{X}(n) \cap B \neq \varnothing\}}{P\{\vec{X}(n) \cap B \neq \varnothing\}}$$

$$= \frac{\sum\limits_{\vec{X} \cap B \neq \varnothing} P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) = \vec{X}\} P\{\vec{X}(n) = \vec{X}\}}{P\{\vec{X}(n) \cap B \neq \varnothing\}}$$

$$\leqslant a_n \frac{\sum\limits_{\vec{X} \cap B \neq \varnothing} P\{\vec{X}(n) = \vec{X}\}}{P\{\vec{X}(n) \cap B \neq \varnothing\}} = a_n$$

同理可证

$$\beta_n^B \leqslant b_n$$

由定理条件及定理 3.2.1 可证。

**定理 3.2.4**　若对于任意满意解集 $B$ 有

$$P\{\vec{X}(n+1) \cap B^c \neq \varnothing / \vec{X}(n) = \vec{X}\} \leqslant \bar{a}_n \quad (\vec{X} \cap B^c = \varnothing)$$

$$P\{\vec{X}(n+1) \cap B^c \neq \varnothing / \vec{X}(n) = \vec{X}\} \leqslant \bar{b}_n \quad (\vec{X} \cap B^c \neq \varnothing)$$

且满足条件：

(1) $\sum\limits_{n=1}^{\infty} (1 - \bar{b}_n) = \infty$

(2) $\bar{a}_n / (1 - \bar{b}_n) \to 0$

则 $\{\vec{X}(n)\}$ 概率强收敛到全局最优解集 $M$。

**证明**　类似定理 3.2.3 可证。

**推论 3.2.1**　若 $\{\vec{X}(n)\}$ 的转移概率满足：

$$P\{\vec{X}(n+1) \cap B \neq \varnothing / \vec{X}(n) = \vec{X}\} = \begin{cases} 1, & \vec{X} \cap B \neq \varnothing \\ \geqslant \delta, & \vec{X} \cap B = \varnothing \end{cases}$$

则 $\{\vec{X}(n)\}$ 概率弱收敛到全局最优解集 $M$。

**证明**　由于 $a_n = 0, b_n = 1 - \delta$，由定理 3.2.3 可证。

**推论 3.2.2**　若 $\{\vec{X}(n)\}$ 的转移概率满足：

$$P\{\vec{X}(n+1) \cap B^c = \varnothing / \vec{X}(n) = \vec{X}\}$$

$$= \begin{cases} 1, & \vec{X} \cap B^c \neq \varnothing \\ \geqslant \delta, & \vec{X} \cap B^c = \varnothing \end{cases}$$

则$\{\vec{X}(n)\}$概率强收敛到全局最优解集 $M$。

**证明**　类似推论 3.2.1,利用定理 3.2.2 可证。

**例 3.2.1**　设遗传算法为 $G = S \circ m$,满足

$$\vec{X}(n+1) = G(\vec{X}(n)) = S \circ m(\vec{X}(n))$$
$$= \begin{cases} m(\vec{X}(n)), & m(\vec{X}(n)) \in B^*(\vec{X}(n)) \\ \vec{X}(n), & m(\vec{X}(n)) \notin B^*(\vec{X}(n)) \end{cases}$$

其中 $S$ 为选择算子,$m$ 为变异算子,且

$$B^*(\vec{X}(n)) = \{\vec{Y}; f(b(\vec{Y})) \geqslant f(b(\vec{X}(n)))\}$$

对于满意解集 $B$,记 $B^* = \{\vec{X}; \vec{X} \cap B \neq \varnothing\}$,则

$$P\{\vec{X}(n+1) \cap B \neq \varnothing / \vec{X}(n) = \vec{X}\}$$
$$= \sum_{\vec{Y}} P_m^n\{\vec{X}, \vec{Y}\}[\chi_{B^*(\vec{X}) \cap B^*}(\vec{Y}) + \chi_{(B^*)^c(\vec{X})}(\vec{Y})\chi_{B^*}(\vec{X})]$$
$$= P_m^n\{\vec{X}, B^* \cap B^*(\vec{X})\} + \chi_{B^*(\vec{X})} P_m\{\vec{X}, B^*(\vec{X})^c\}$$

由于 $B^*(\vec{X})$ 与 $B^*$ 均为满意解集,于是当 $\vec{X} \cap B = \varnothing$ 时有 $b(\vec{X}) \notin B$。由于对 $\forall \vec{Y} \in B^*$ 有 $b(\vec{Y}) \in B$,故 $f(b(\vec{Y})) \geqslant f(b(\vec{x}))$,即 $\vec{Y} \in B^*(\vec{X})$,于是 $B^* \subset B^*(\vec{X})$,$B^* \cap B^*(\vec{X}) = B^*$,因此

$$P\{\vec{X}(n+1) \cap B \neq \varnothing / \vec{X}(n) = \vec{X}\} = P_m^n\{\vec{X}, B^*\}$$

当 $\vec{X} \cap B \neq \varnothing$ 时 $B^*(\vec{X}(n)) \subset B^*$,$B^*(\vec{X}) \cap B^* = B^*(\vec{X})$,于是

$$P\{\vec{X}(n+1) \cap B \neq \varnothing / \vec{X}(n) = \vec{X}\}$$
$$= P_m^n\{\vec{X}, B^*(\vec{X})\} + P_m^n\{\vec{X}, B^*(\vec{X})^c\} = 1$$

当 $\vec{X} \cap B = \varnothing$ 时,若有 $P_m^n(\vec{X}, B^*) \geqslant \delta > 0$,由推论 3.2.1 可证 $\{\vec{X}(n)\}$ 概率弱收敛到全局最优解集 $M$。特别取均匀随机变异算子,当 $\vec{X} \cap B = \varnothing$ 时

$$P_m^n\{\vec{X}, B^*\} \geqslant p^l = \delta > 0$$

则满足所需条件。

**例 3.2.2**　采用与例 3.2.1 相同的遗传算法，取代 $B^*(\vec{X}(n))$ 为

$$\tilde{B}(\vec{X}(n)) = \{\vec{Y}; f(d(\vec{Y})) \geqslant f(d(\vec{X}(n)))\}$$

则可以得到

$$P\{\vec{X}(n+1) \cap B^c = \varnothing / \vec{X}(n) = \vec{X}\}$$

$$= \begin{cases} 1, & \vec{X} \cap B^c = \varnothing \\ P_m^n(\vec{X}, B^*), & \vec{X} \cap B^c \neq \varnothing \end{cases}$$

只要 $P_m^n(\vec{X}, B^*) \geqslant \delta$，则 $\{\vec{X}(n)\}$ 概率强收敛到全局最优解集 $M$。

记

$$d_n = \min\{P_m^n(X, Y);\ X \neq Y,\ X, Y \in S\}$$

$$r_n = \min\{P_m^n(X, Y);\ X, Y \in S\}$$

有以下定理。

**定理 3.2.5**　对于例 3.2.1，若 $\sum_{n=1}^{\infty} d_n = \infty$，则 $\vec{X}(n) \to M(\text{P.W.})$。对于例 3.2.2，若 $\sum_{n=1}^{\infty} r_n^N = \infty$，则 $\vec{X}(n) \to M(\text{P.S.})$。

**证明**　当 $\vec{X} \cap B = \varnothing$ 时，有

$$P\{\vec{X}(n+1) \cap B \neq \varnothing / \vec{X}(n) = \vec{X}\}$$

$$= P_m^n\{\vec{X}, \vec{Y} \cap B \neq \varnothing\}$$

$$\geqslant \sum_{Y_1 \in B} \prod_{j=1}^{N} P_m^n\{X_j, Y_j\} \geqslant d_n$$

结合例 3.2.1 与定理 3.2.3 则证 $\vec{X}(n) \to M(\text{P.W.})$。

当 $\vec{X} \cap B^c \neq \varnothing$ 时，有

$$P\{\vec{X}(n+1) \cap B^c = \varnothing / \vec{X}(n) = \vec{X}\}$$

$$= P_m^n\{\vec{X}, \vec{Y} \cap B^c = \varnothing\} \geqslant r_n^N$$

结合例 3.2.2 与定理 3.2.4 则证 $\vec{X}(n) \to M(\text{P.S.})$。

**例 3.2.3**(父代参与竞争无杂交遗传算法)　父代参与竞争无

杂交遗传算法为

$$\vec{X}(n+1)=S\circ m(\vec{X}(n))$$

其中 $m$ 为变异算子，$S$ 为选择算子，它的转移概率为

$$P\{\vec{X}(n+1)=\vec{Z}/\vec{X}(n)=\vec{X}\}$$
$$=\sum_{\vec{Y}}P_m^n\{\vec{X},\vec{Y}\}P_s^n\{\langle\vec{X}\oplus\vec{Y}\rangle,\vec{Z}\} \tag{3.2.5}$$

其中 $P_m^n$ 为变异算子转移概率，$P_s^n$ 为选择算子转移概率，且选择算子满足以下条件：

(1) $\langle\vec{X}\oplus\vec{Y}\rangle\subset\vec{X}\cup\vec{Y}$，$\langle\vec{X}\oplus\vec{Y}\rangle\in S^N$

(2) 对于任意 $X\in\langle\vec{X}\oplus\vec{Y}\rangle$，$Y\notin\langle\vec{X}\oplus\vec{Y}\rangle$，$f(X)\geqslant f(Y)$

(3) 选择算子的转移概率为：

$$P_s^n\{\langle\vec{X}\oplus\vec{Y}\rangle,\vec{Z}\}$$
$$=\prod_{k=1}^{N}\left[f(Z_k)/\sum_{X\in\langle\vec{X}\oplus\vec{Y}\rangle}f(X)\right]\quad(\vec{Z}\subset\langle\vec{X}\oplus\vec{Y}\rangle) \tag{3.2.6}$$

下面对于父代参与竞争无杂交的遗传算法给出概率收敛定理。

**定理 3.2.6** 对于父代参与竞争无杂交的遗传算法，若

$$\sum_{n=1}^{\infty}r_n^N=\infty$$

则 $\{\vec{X}(n)\}$ 概率强收敛到全局最优解集 $M$，其中

$$r_n=\min\{P_m^n\{X,Y\};X,Y\in S\}$$

**证明** 对于任意满意解集 $B$，当 $\vec{X}\cap B^c=\varnothing$ 时，对于任意 $\vec{Y}$，$\langle\vec{X}\oplus\vec{Y}\rangle\cap B^c=\varnothing$，于是

$$P_s^n\{\langle\vec{X}\oplus\vec{Y}\rangle,\vec{X}(n+1)\cap B^c\neq\varnothing\}=0$$

从而 $\bar{a}_n=0$。当 $\vec{X}\cap B^c\neq\varnothing$ 时取 $\vec{Y}\cap B^c=\varnothing$，$\langle\vec{X}\oplus\vec{Y}\rangle\cap B^c=\varnothing$，于是

$$P_s^n\{\langle\vec{X}\oplus\vec{Y}\rangle,\vec{X}(n+1)\cap B^c=\varnothing\}=1$$

从而

$$\sum_{\vec{Y}} P_m^n\{\vec{X},\vec{Y}\} P_s^n\{\langle \vec{X} \oplus \vec{Y}\rangle, \vec{X}(n+1) \cap B^c = \varnothing\}$$

$$\geqslant \sum_{\vec{Y}\cap B^c=\varnothing} P_m^n\{X,Y\} = P_m^n\{X,\vec{Y} \cup B^c = \varnothing\} \geqslant r_n^N$$

则 $b_n = 1 - r_n^N$。由定理 3.2.4 则证。

**例 3.2.4**(最佳值选择遗传算法)　最佳值选择遗传算法为：

$$\vec{X}(n+1) = C \circ S \circ m(\vec{X}(n))$$

其中 $C$ 为杂交算子，$S$ 为选择算子，$m$ 为变异算子，转移概率为

$$P\{\vec{X}(n+1) = \vec{Z}/\vec{X}(n) = \vec{X}\}$$

$$= \sum_{\vec{Y}} P_m^n\{\vec{X},\vec{Y}\} \sum_{V_j, W_j \in B(\vec{Y})} \prod_{1}^{N} P_s^n\{\vec{Y}, V_j \otimes W_j\} \cdot P_c^n(V_j \otimes W_j, Z_j) \qquad (3.2.7)$$

其中 $P_m^n$ 为第 $n$ 步变异转移概率，$P_s^n$ 为第 $n$ 步选择转移概率，$P_c^n$ 为第 $n$ 步杂交转移概率。记

$$B(\vec{Y}) = \{Y_i; f(Y_i) \geqslant f(Y_j)\ (j \leqslant N)\}$$

$$\vec{Y}(X) = \{Y_i; Y_i = X\}$$

选择算子转移概率为

$$P_s^n\{\vec{Y}, V \otimes W\} = \frac{|\vec{Y}(V)|\,|\vec{Y}(W)|}{|B(\vec{Y})|^2} \quad (V, W \in B(\vec{Y}))$$

下面给出最佳值选择遗传算法的概率收敛定理。

记

$$a_n = \min\{P_m^n(X,X);\ X \in S\}$$

$$d_n = \min\{P_m^n(X,Y);\ X \neq Y\}$$

$$b_n = \max\{P_m^n(X,Y);\ X \neq Y\}$$

$$\bar{a}_n = \min\{P_c^n(X \otimes Y, X);\ X, Y \in S\}$$

**定理 3.2.7**　对于最佳值选择遗传算法，$\{\vec{X}(n)\}$ 概率强收敛到全局最优解集的条件是：

(1) $\sum\limits_{n=1}^{\infty}(1-\bar{a}_n)<\infty$

(2) $\sum\limits_{n=1}^{\infty}d_n=\infty$, $\lim\limits_{n\to+\infty}a_n^{N-1}=a_n>0$

(3) $\lim\limits_{n\to+\infty}\dfrac{b_n^N}{d_n}=0$, $\lim\limits_{n\to+\infty}\dfrac{1-\bar{a}_n}{d_n}=0$

**证明** 记

$$P\{\vec{X},\vec{Y},\vec{X}(n+1)\cap B^c\neq\varnothing\}$$

$$=P_m^n\{\vec{X},\vec{Y}\}\sum_{V_j,W_j\in B(\vec{Y})}\prod_{j=1}^{N}P_s^n\{\vec{Y},V_j\otimes W_j\}$$

$$\cdot P_c^n\{V_j\otimes W_j,\vec{X}(n+1)\cap B^c\neq\varnothing\}$$

当 $\vec{Y}\cap B\neq\varnothing$ 时，$V_j,W_j\in B$，则

$$P_c^n\{V_j\otimes W_j,Z_j\notin B\}$$

$$\leqslant 1-P_c^n\{V_j\otimes W_j,V_j\}\leqslant 1-\bar{a}_n$$

当 $\vec{Y}\cap B=\varnothing$ 时，$V_j,W_j\notin B$，于是

$$\sum_{\vec{Y}\cap B=\varnothing}P\{\vec{X},\vec{Y},\vec{X}(n+1)\cap B^c\neq\varnothing\}\leqslant\sum_{\vec{Y}\cap B=\varnothing}P_m^n\{\vec{X},\vec{Y}\}$$

$$\leqslant P_m^n\{\vec{X},\vec{Y}\cap B=\varnothing\}$$

分两种情形考虑：

(1) 当 $\vec{X}\cap B^c=\varnothing$ 时，与 $\vec{Y}\cap B=\varnothing$ 的 $\vec{Y}$ 总有$X_i\neq Y_i$($i\leqslant N$)，于是

$$P_m^n\{\vec{X},\vec{Y}\cap B=\varnothing\}\leqslant b_n^N|S^N|$$

(2) 若 $\vec{X}\cap B^c\neq\varnothing$ 时

$$P_m^n\{\vec{X},\vec{Y}\cap B=\varnothing\}=1-P_m^n\{\vec{X},\vec{Y}\cap B\neq\varnothing\}$$

$$\leqslant 1-d_n\cdot a_n^{N-1}$$

于是 $\vec{X}\cap B^c=\varnothing$ 时

$$P_m^n\{\vec{X}(n+1)\cap B^c\neq\varnothing/\vec{X}(n)=\vec{X}\}$$

$$\leqslant(1-\bar{a}_n)+b_n^N|S^N|=a_n$$

当 $\vec{X} \cap B^c \neq \varnothing$ 时

$$P_m^n\{\vec{X}(n+1) \cap B^c \neq \varnothing / \vec{X}(n) = \vec{X}\}$$
$$\leqslant (1-\bar{a}_n) + (1-d_n a_n^{N-1}) = b_n$$

根据定理条件即得

$$\sum_{n=1}^{\infty}(1-b_n) = \sum_{n=1}^{\infty}[d_n a_n^{N-1} - (1-\bar{a}_n)]$$
$$= \sum_{n=1}^{\infty} d_n a_n^{N-1} - \sum_{n=1}^{\infty}(1-\bar{a}_n) = \infty$$

$$\lim_{n\to\infty}\frac{a_n}{1-b_n} = \lim_{n\to\infty}\frac{(1-\bar{a}_n) + b_n^N |S^N|}{d_n a_n^{N-1} - (1-\bar{a}_n)}$$
$$= |S^N| \lim_{n\to\infty}\frac{b_n^N}{d_n} = 0$$

根据定理 3.2.4 即证

$$\lim_{n\to\infty} P\{\vec{X}(n) \cap B^c = \varnothing\} = 1$$

**例 3.2.5**　对于最佳值选择遗传算法，取

$$P_m^n(X,Y) = p_n^{d(X,Y)}(1-p_n)^{l-d(X,Y)}$$

其中

$$d(X,Y) = |\{k;\ X_k \neq Y_k(k \leqslant l)\}|$$

我们有

$$b_n = (1-p_n)^{l-1}p_n,\quad d_n = p_n^l,\quad a_n = (1-p_n)$$

对于单点杂交

$$P_c^n\{X \otimes Y, Z\} = \begin{cases} hq_n/l, & Z \neq X \\ (1-q_n) + hq_n/l, & Z = X \end{cases}$$

有

$$\bar{a}_n = 1 - q_n$$

对于均匀杂交

$$P_c^n\{X \otimes Y, Z\} = \prod_{i=1}^{l}[(1-q_n)(X_i \oplus Z_i) + q_n(Y_i \oplus Z_i)]$$

其中 $a \oplus b = 1$ 当且仅当 $a = b$,同样有

$$\bar{a}_n = 1 - q_n$$

根据定理 3.2.7,当满足条件

(1) $\lim\limits_{n\to\infty} p_n = 0$

(2) $N > l$

(3) $\sum\limits_{n=1}^{\infty} q_n < \infty$

(4) $\sum\limits_{n=1}^{\infty} p_n^l = \infty$

时,$\{\vec{X}(n)\}$ 概率强收敛到全局最优解集。特别取 $p_n = \sqrt[l]{\frac{1}{n}}$,则 $q_n = \frac{1}{n^2}$ 满足上述条件。

对于父代参与竞争有杂交的遗传算法可以表示为模型

$$\vec{X}(n+1) = G_n(\vec{X}(n)) = S_n(\vec{X}(n), m_n \circ C_n(\vec{X}(n))$$

其中 $C_n$ 为杂交算子。$C_n$ 为$S^N \to S^{N_0}$ 的随机算子,$m_n$ 为$S^N \to S^{N_0}$ 的随机算子,$S_n$ 为 $S^{N+N_0} \to S^N$ 的随机选择算子。

**例 3.2.6**(整体退火遗传算法) $\{\vec{X}(n)\}$ 的转移概率为

$$P\{\vec{X}(n+1) = \vec{Y}/\vec{X}(n) = \vec{X}\}$$
$$= \sum_{\vec{W},\vec{Z}\in S^{N_0}} P_c^n\{\vec{X},\vec{W}\}P_m^n\{\vec{W},\vec{Z}\}P_s^n\{\vec{X}\oplus\vec{Z},\vec{Y}\}$$

其中

$$P_s^n\{\vec{X}\oplus\vec{Z},\vec{Y}\}$$
$$= \prod_{j=1}^{N}\left[e^{f(Y_j)/T_n}\Big/\left(\sum_{i=1}^{N}e^{f(X_i)/T_n} + \sum_{i=1}^{N_0}e^{f(Z_i)/T_n}\right)\right]$$

**定理 3.2.8** 整体退火遗传算法,当 $T_n \to 0$ 时,$\{\vec{X}(n)\}$ 概率强收敛到最优解集 $M$,即

$$\lim_{n\to\infty} P\{\vec{X}(n) \subset M\} = 1$$

**证明**　当 $\vec{X} \cap M^c = \varnothing, \vec{X} \subset M$ 时，对于任意 $\vec{Z} \in S^{N_0}$ 有

$$\max\{f(X); X \in \vec{X} \oplus \vec{Z}\} = f^*$$

若 $\vec{Y} \cap M^c \neq \varnothing$，必存在 $Y_{i_0} \in \vec{Y}$，使 $Y_{i_0} < f^*$，于是

$$\begin{aligned} & P_S^n\{(\vec{X} \oplus \vec{Z}, \vec{Y}\} \\ & \leqslant e^{f(Y_{i_0})/T_n} / \left[ N e^{f^*/T_n} + \sum_{j=1}^{N_0} e^{f(Z_j)/T_n} \right] \\ & \leqslant e^{(f(Y_{i_0}) - f^*)/T_n} \leqslant e^{-\delta/T_n} \end{aligned}$$

其中

$$\delta = \min\{|f(X) - f(Y)|; f(X) \neq f(Y)\}$$

于是

$$\begin{aligned} & P\{\vec{X}(n+1) \cap M^c \neq \varnothing / \vec{X}(n) = \vec{X}\} \\ & \quad = \sum_{\vec{Y} \cap M^c \neq \varnothing} \sum_{\vec{Z} \in S^{N_0}} \sum_{\vec{W} \in S^{N_0}} P_m^n\{\vec{X}, \vec{W}\} \\ & \quad \cdot P_m^n\{\vec{W}, \vec{Z}\} P_{S_n}\{\vec{X} \oplus \vec{Z}, \vec{Y}\} \\ & \quad \leqslant |S^N| e^{-\delta/T_n} = \bar{a}_n \end{aligned}$$

当 $\vec{X} \cap M^c \neq \varnothing, \vec{Y} \cap M^c = \varnothing$ 时，$\vec{Y} \subset M$，则

$$\begin{aligned} & P_{S_n}\{\vec{X} \oplus \vec{Z}, \vec{Y}\} \\ & \quad \geqslant \prod_{j=1}^{N} \left[ e^{f^*/T_n} \Big/ \sum_{j=1}^{N+N_0} e^{f^*/T_n} \right] \\ & \quad = \left( \frac{1}{N+N_0} \right)^N \end{aligned}$$

从而

$$\begin{aligned} & P\{\vec{X}(n+1) = \vec{Y} / \vec{X}(n) = \vec{X}\} \\ & \quad \geqslant \left( \frac{1}{N+N_0} \right)^N \end{aligned}$$

于是当 $\vec{X} \cap M^c \neq \varnothing$ 时

$$P^n\{\vec{X}, \vec{Y} \cap M^c \neq \varnothing\}$$

$$\leqslant 1 - P^n\{\vec{X}, \vec{Y} \cap M^c = \varnothing\}$$

$$\leqslant 1 - \left(\frac{1}{N + N_0}\right)^N = \bar{b}_n$$

根据定理 3.2.4 则证。

**定理 3.2.9** 在整体退火遗传算法中，若父代不参加竞争，则 $\{\vec{X}(n)\}$ 不可能概率强收敛到最优解集 $M$，即

$$\lim_{n\to\infty} P\{\vec{X}(n) \subset M\} < 1$$

**证明** 取 $X \in S$，且 $f(X) = f < f^*$，对于

$$\vec{X}(n+1) = (X, \cdots, X) = \vec{Y}$$

存在 $\vec{Z}$，使 $\vec{X}(n+1) \subset \vec{Z}$，于是

$$P_s^n\{\vec{Z}, \vec{Y}\}$$

$$= \prod_{b\in \vec{Y}} \left[ e^{f(b)/T_n} / \sum_{b\in \vec{Z}} e^{f(b)/T_n} \right]$$

$$= \left[ e^{f(b)/T_n} / \sum_{b\in \vec{Z}} e^{f(b)/T_n} \right]^N$$

$$\geqslant \left(\frac{1}{N_0}\right)^N$$

从而，对于任意 $\vec{X} \in S^N$

$$P\{\vec{X}(n+1) = \vec{Y} / \vec{X}(n) = \vec{X}\}$$

$$\geqslant \left(\frac{1}{N_0}\right)^N$$

$$P\{\vec{X}(n+1)\}$$

$$= \sum_{\vec{X}\in S^N} P\{\vec{X}(n+1) = \vec{Y} / \vec{X}(n) = \vec{X}\} P\{\vec{X}(n) = \vec{X}\}$$

$$\geqslant \left(\frac{1}{N_0}\right)^N$$

$$P\{\vec{X}(n+1) \subset M\}$$

$$\leqslant 1 - P\{\vec{X}(n+1) = \vec{Y}\}$$

$$\leqslant 1 - \left(\frac{1}{N_0}\right)^N$$

则证。

## 3-3　抽象遗传算法的概率收敛定理

在 3-2 节中我们看到 $\alpha_n^B, \beta_n^B$ 或 $\bar{\alpha}_n^B, \bar{\beta}_n^B$ 在讨论遗传算法概率收敛中是非常重要的。本节将进一步针对抽象遗传算法给出概率收敛定理。

设 $\vec{X}$ 和 $\vec{Y}$ 为种群，记

$$B(\vec{X}) = \{X_i;\ f(X_i) \geqslant f(X_j)\ (j \leqslant N)\}$$

$$\hat{f}(\vec{X}) = \max\{f(X_i);\ X_i \in \vec{X}\}$$

则 $B(\vec{X})$ 是 $\vec{X}$ 中适应值最大的个体的全体，$\hat{f}(\vec{X})$ 是 $\vec{X}$ 中的个体的最大适应值。记

$$\vec{Y}(\vec{X}) = \{Y_i \in \vec{Y};\ f(Y_i) = \hat{f}(\vec{X})\}$$

则 $\vec{Y}(\vec{X})$ 表示 $\vec{Y}$ 种群中那些适应值达到 $\vec{X}$ 种群中最大适应值的个体。用 $|A|$ 表示 $A$ 中元素个数。

**定义 3.3.1**(抽象选择算子)　选择算子 $S$ 是一个随机映射 $S: S^N \rightarrow S^N$，满足：

(1) 对于任意 $\vec{X} \in S^N, S(\vec{X}) \subset \vec{X}$；

(2) 对于任意 $\vec{X} \in S^N, |B(\vec{X})| < N$ 有

$$P\{|S(\vec{X})(\vec{X})| > |B(\vec{X})|\} > 0 \tag{3.3.1}$$

上述两条描述了选择算子的一般特征：(1) 选择是在原种群中选择；(2) 选择结果要有可能增加种群中最大适应值个体的数目，只要当前种群不是齐次种群。

**例 3.2.1**　假定选择为比例选择，即

$$P\{S(\vec{X})_i = X_j\} = f(X_j) / \sum_{m=1}^{N} f(X_m)$$

不妨假定 $|B(\vec{X})| = N - 1, X_N \notin B(\vec{X})$，于是

$$P\{|S(\vec{X})(\vec{X})| > |B(\vec{X})|\}$$

$$= P\{S(\vec{X})_i \neq X_N (i \leqslant N)\} = \left(1 - \frac{f(X_N)}{\sum_{j=1}^{N} f(X_j)}\right)^N > 0$$

于是满足定义 3.3.1 的(1) 和(2)。

**定义 3.3.2**(抽象遗传算子) 抽象遗传算子 $G$ 是一个随机映射 $G: S^N \to S^N$,它对于任何满意解集 $B$ 均满足

(1) $\alpha^B = P\{G(\vec{X}) \cap B = \varnothing / \vec{X} \cap B \neq \varnothing\} < 1$

(2) $\beta^B = P\{G(\vec{X}) \cap B = \varnothing / \vec{X} \cap B = \varnothing\} < 1$

**例3.3.2** 假定 $G$ 是均匀变异算子,变异率为 $p(0 < p < 1)$,设 $B$ 为满意解集,取 $\vec{X}$ 使 $\vec{X} \cap B = \varnothing$,即 $X_i \notin B(i \leqslant N)$。取 $Y \in B, \vec{Y} = (X_1, \cdots, X_{N-1}, Y)$,则 $\vec{Y} \cap B \neq \varnothing$,且

$$P\{G(\vec{X}) = \vec{Y} / \vec{X}\}$$

$$= (1-p)^{l(N-1)} p^{d(X_N, Y)} (1-p)^{l - d(X_N, Y)} > 0$$

于是

$$P\{G(\vec{X}) \cap B \neq \varnothing / \vec{X} \cap B = \varnothing\} > 0$$

从而

$$\alpha^B = P\{G(\vec{X}) \cap B = \varnothing / \vec{X} \cap B = \varnothing\} < 1$$

同理可证

$$\beta^B = P\{G(\vec{X}) \cap B = \varnothing / \vec{X} \cap B = \varnothing\} < 1$$

因而均匀变异算子满足定义 3.3.2。

我们指出,由于杂交算子在模式 $\mathscr{L}(\vec{X})$ 中进行,当 $B \cap \mathscr{L}(\vec{X}) = \varnothing$ 时,必有

$$\beta^B = P\{G(\vec{X}) \cap B = \varnothing / \vec{X} \cap B = \varnothing\} = 1$$

但是,当 $\vec{X} \cap B \neq \varnothing$ 时,$B \cap \mathscr{L}(\vec{X}) \neq \varnothing$,于是

$$\alpha^B = P\{G(\vec{X}) \cap B = \varnothing / \vec{X} \cap B \neq \varnothing\} < 1$$

因此,仅用杂交算子不构成遗传算子。

**定义 3.3.3**(抽象遗传算法) 一个抽象遗传算法是由一列选

择算子$\{S_n\}$和一列遗传算子$\{G_n\}$复合构成的$S^N$上的随机过程，并表示为下列的状态方程：

$$\vec{X}(n+1) = G_n \circ S_n(\vec{X}(n)) \tag{3.3.2}$$

其中$\vec{X}(0) \in S^N$为随机选取的初始种群。转移概率为

$$P\{\vec{X}(n+1) = \vec{Y}/\vec{X}(n) = \vec{X}\}$$
$$= \sum_{\vec{Z}} P_{S_n}(\vec{X},\vec{Z}) P_{G_n}(\vec{Z},\vec{Y}) \tag{3.3.3}$$

其中$P_{S_n}$为$S_n$的转移概率，$P_{G_n}$为$G_n$的转移概率。

由于给定$\vec{X}(n)$后，通过算子唯一确定了$\vec{X}_{n+1}$的分布，与$\vec{X}_0,\cdots,\vec{X}_{(n+1)}$的状态无关，则$\{\vec{X}(n)\}$是马尔可夫链。特别当$S_n$，$G_n$与$n$无关时，$\{\vec{X}(n)\}$为齐次马尔可夫链。

**定义 3.3.4**(选择压与选择强度)　设$S$为抽象选择算子，称

$$P_s = \sup\{\rho \in R^+; P\{|S(\vec{X})(\hat{X})| \geqslant \rho + 1\} > 0,$$
$$|B(\vec{X})| = 1\} \tag{3.3.4}$$

为选择压，而称

$$I_s = \inf\{P\{|S(\vec{X})(\hat{X})| \geqslant P_s + 1\}; |B(\vec{X})| = 1)\} \tag{3.3.5}$$

为选择强度。

**例 3.3.3**(比例选择)　设$\sigma$为正单调增函数，对于任意种群$\vec{X} \in S^N$，选择概率为

$$P\{S(\vec{X})_i = X_j\} = \sigma(f(X_j)) / \sum_{m=1}^{N} \sigma(f(X_m))$$

记

$$\rho' = \max\left\{\frac{\sigma(f(X))}{\sigma(f(Y))}, \sigma(f(X)) < \sigma(f(Y)), X, Y \in S\right\}$$

则有选择压与选择强度。

$$P_s = N - 1,\ I_s = (1 + (N-1)\rho')^{-N}$$

由于$|B(\vec{X})| = 1$，不妨假定$X_i \in B(\vec{X})$，则

$$P\{|G(\vec{X})(\hat{X})| \geqslant N\}$$
$$= P\{G(\vec{X})_j = X_i (j \leqslant N)\} = \alpha_i^N > 0$$

其中

$$\alpha_i = \sigma(f(X_i)) / \sum_{j=1}^{N} \sigma(f(X_j))$$

则证 $P_s = N - 1$。又因为

$$\alpha_i \geqslant \left[1 + \sum_{j \neq i}^{N} \frac{\sigma(f(X_j))}{\sigma(f(X_i))}\right]^{-1} \geqslant [1 + (N-1)\rho']^{-1}$$

从而

$$I_s = [1 + (N-1)\rho']^{-N}$$

**例 3.3.4**(线性 ranking 选择)　假定种群 $\vec{X}$ 满足 $i < j$ 时 $f(X_i) > f(X_j)$,则选择算子为

$$P\{S(\vec{X})_i = X_j\} = p + (N - j)q$$

其中 $p,q$ 满足

$$\frac{(N-1)N}{2}q + Np = 1$$

则

$$P_s = N - 1,\ I_s = \left(\frac{1}{N} + \frac{N-1}{2}q\right)^N$$

**例 3.3.5**(非线性 ranking 选择)　假定种群 $\vec{X}$ 满足 $i < j$ 时 $f(X_i) > f(X_j)$,则选择算子为

$$P\{S(\vec{X})_i = X_j\} = q(1-q)^{j-1}$$

其中 $q$ 满足

$$\frac{q(1-q^N)}{1-q} = 1$$

则

$$P_s = N - 1,\qquad I_s = 1$$

**例 3.3.6**(tournament 选择)　$S(\vec{X})_i$ 是取 $X_i$ 相邻的 $s$ 个个体

中的最佳个体，则

$$P_s = s-1, \qquad I_s = 1$$

**例 3.3.7**(boltzman 选择)　$S(\vec{X})_i$ 是从 $X_i$ 相邻的两个个体中确定。若 $f(X_{i+1}) \geqslant f(X_{i-1})$，则

$$P\{S(\vec{X})_i = f(X_{i+1})\}$$
$$= 1\Big/\left(1+\exp\left(\frac{f(X_{i-1})-f(X_{i+1})}{T}\right)\right)$$

取

$$\Delta = \min\{f(X)-f(Y);\ f(X)>f(Y)\}$$

则有

$$P_s = 1, \qquad I_s = \left(1+\exp\left(-\frac{\Delta}{T}\right)\right)^{-2}$$

**例 3.3.8**(elitist 选择)　将 $\vec{X}$ 中的 $d(\vec{X})$ 代替为 $b(\vec{X})$，其它不变，则

$$P_s = 1, \qquad I_s = 1$$

**定义 3.3.5**(遗传吸收率与散射率)　设 $G$ 为抽象遗传算子，称

$$A_G = \inf\{P[\vec{X},G(\vec{X})\cap B \neq \varnothing];\vec{X}\cap B = \varnothing,\ B\subset S\} \tag{3.3.6}$$

为遗传吸收率。称

$$S_G^r = \sup\{P[\vec{X},G(\vec{X})\cap B = \varnothing];|\vec{X}\cap B| \geqslant r\ B\subset S\} \tag{3.3.7}$$

为遗传散射率。

$A_G = \inf\{1-\beta^{l};\vec{X}\cap B = \varnothing, B\subset S\}$ 表示吸收能力；$S_G^r$ 表示散射能力。为了保证遗传算法收敛效果，一般要求 $A_G$ 较大，$S_G^r$ 较小。

**例 3.3.9**(变异算子)　对于种群 $\vec{X}\in S^N$，变异算子 $m$ 的转移概率为

$$P\{m(\vec{X})_{ij} = 1 - x_{ij}\} = p_m$$
$$P\{m(\vec{X})_{ij} = x_{ij}\} = 1 - p_m$$

这里 $p_m$ 表示变异概率，$x_{ij}$ 表示$X_i$ 的第$j$ 个分量，取值0或1。如果 $p_m \leqslant 1/2$，则有

$$A_G = 1 - (1 - p_m^l)^N$$
$$S_G^r = [1 - (1 - p_m)^l]^r (1 - p_m^l)^{N-r}$$

下面，证明这一事实。按照 $A_G$ 定义，有

$$A_G = \inf\{P\{\vec{X}, m(\vec{X}) \cap B \neq \varnothing\}; \vec{X} \cap B = \varnothing,\\ B \subset S\}$$

显然 inf 可以在下述情况下达到：

$$B = \{(0,0,\cdots,0)\}, \qquad \vec{X} = \{(X_1, X_2, \cdots, X_N)\}$$

其中

$$X_i = (1,1,\cdots,1)$$

这时条件 $\vec{X} \cap B = \varnothing$ 成立。要使 $m(\vec{X}) \cap B \neq \varnothing$，至少应有一个 $X_i$ 变为$(0,0,\cdots,0)$。$X_i$ 变为$(0,0,\cdots,0)$ 的概率为 $p_m^l$，因此至少有一个 $X_i$ 变为$(0,0,\cdots,0)$ 的概率为

$$A_G = 1 - (1 - p_m^l)^N$$

下面进一步计算

$$S_G^r = \sup\{P\{\vec{X}, m(\vec{X}) \cap B = \varnothing\}; |\vec{X} \cap B| \geqslant r, B \subset S\}$$

不妨假定$|\vec{X} \cap B| = r_1 \geqslant r$，这时在以下的条件下使 $S_G^r$ 中sup达到：

(1) $B = \{(0,0,\cdots,0)\}$

(2) $\vec{X}$ 中有 $r_1$ 个个体为$(0,0,\cdots,0)$，$(N - r_1)$ 个个体为$(1,1,\cdots,1)$

在上述条件下$|B \cap \vec{X}| = r_1$，要使 $m(\vec{X}) \cap B = \varnothing$，必须满足以下两个条件：

(1) 每个个体 $X_i = (0,0,\cdots,0)$ 至少有一位变异到1；

(2) 每个个体 $X_j = (1,1,\cdots,1)$ 不变异到$(0,0,\cdots,0)$，于是

$$S_G^r \leqslant [1-(1-p_m)^l]^{r_1}(1-p_m^l)^{N-r_1}$$

当 $p_m \leqslant 1/2$ 时，上式右端关于 $p_m$ 单调减，从而

$$S_G^r \leqslant [1-(1-p_m)^l]^r(1-p_m^l)^{N-r}$$

**例 3.3.10**(组合算子)　组合算子是杂交算子 $C$ 与变异算子 $m$ 的复合运算，即

$$G = m \circ C$$

设 $\vec{Y} = C(\vec{X}), \vec{Z} = m(\vec{Y})$。

对于杂交算子 $C$ 产生的 $\vec{Y} = (Y_1,\cdots,Y_N)$ 由 $N$ 次独立实验产生。对于 $i \leqslant N$, $Y_i$ 产生过程如下：首先从 $\vec{X}$ 中选择一对母体 $Y_i', Y_i''$，则

$$P\{Y_i = Y_i' \text{或} Y''_i\} = 1-p_c,\ P\{Y_i = C(Y_i', Y''_i)\} = p_c$$

$C(Y_i', Y''_i)$ 表示通过 $Y_i', Y''_i$ 杂交生成。这时

$$A_G = 1-(1-p_m^l)^N, \qquad S_G^N = [1-(1-p_m)^l]^N$$

对于 $A_G$ 的计算，我们只须考虑齐次种群。这时杂交算子 $C$ 不起作用，由例 3.3.9 可知

$$A_G = 1-(1-p_m^l)^N$$

对于 $S_G^r$ 的计算就要麻烦得多。首先计算 $S_G^N$，这时 $|\vec{X} \cap B| = N$，不妨设

$$B = \{(0,\cdots,0)\}$$
$$\vec{X} = \{(0,0,\cdots,0)_i;\ i \leqslant N\}$$

于是 $\vec{X}$ 为齐次种群，杂交算子不起作用。为使 $\vec{X} \cap B = \varnothing$，必须将 $\vec{X}$ 中每个$(0,0,\cdots,0)$ 进行变异，从而有

$$S_G^N = [1-(1-p_m)^l]^N$$

进一步计算 $S_G^{N-1}$。不妨假定

$$B = \{(0,0,\cdots,0)\}$$
$$\vec{X} = \{(0,\cdots,0),\cdots,(0,\cdots,0),(1,\cdots,1)\}$$

首先考虑杂交 $C(\vec{X})$,有$[(l-1)/l]^2$ 的概率取一母体为(0,0,…,0)和(0,…,0),这时杂交不起作用。有$(2l-1)/l^2$ 的概率取的母体中含 1 个个体(1,…,1)。这时杂交以后的个体分量可能含有 1。于是

$$1-(1-p_m)^l \leqslant \{m \circ C(\vec{X})_i \notin B\} \leqslant \left(\frac{l-1}{l}\right)^2[1-(1-p_m)^l]+\frac{2l-1}{l^2}(1-p_m^l)$$

从而

$$S_G^N \leqslant S_G^{N-1} \leqslant \left\{\left(\frac{l-1}{l}\right)^2[1-(1-p_m)^l] + \frac{2l-1}{l^2}(1-p_m^l)\right\}^N$$

由于 $S_G^r(r<N)$ 计算的复杂性,我们这里不再具体计算。

由上面的分析也可以看到,选择和变异在遗传算法中起着关键作用。从某种意义上讲,杂交的作用不甚明显。

**定理 3.3.1** 假定抽象遗传算法满足以下条件:

(1) $S_n$ 的选择压一致有下界,即存在正整数 $m>0$,使对于任意 $n$ 有 $P_{S_n} \geqslant m$

(2) $\sum_{n=1}^{\infty} A_{G_n} = \infty$

(3) 对于任意 $\vec{X} \in S^N$ 及 $m \geqslant 1$ 有

$$P\{|S(\vec{X})(\hat{X})| \geqslant m\} \geqslant \inf\{P\{|S(\vec{Y})(\hat{Y})| \geqslant m\}; |B(\vec{Y})| = 1\}$$

(4) 对于选择强度 $I_{S_n}$,吸收率 $A_{G_n}$,散射率 $S_{G_n}^{m+1}$ 有关系:

$$\lim_{n\to\infty} \frac{1-I_{S_n}(1-S_{G_n}^{m+1})}{A_{G_n}} = 0$$

则抽象遗传算法概率弱收敛到全局最优解集,即

$$\lim_{n\to\infty} P\{\vec{X}(n) \cap M \neq \varnothing\} = 1$$

**证明**　对于任意满意解集 $B$,记

$$\alpha_n^B = P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) \cap B \neq \varnothing\}$$

$$\beta_n^B = P\{\vec{X}(n+1) \cap B = \varnothing / \vec{X}(n) \cap B = \varnothing\}$$

$$\vec{Y} = S_n(\vec{X}(n))$$

由选择算子定义知 $\vec{Y} \subset \vec{X}(n)$,于是 $\vec{X}(n) \cap B = \varnothing$ 时必有 $\vec{Y} \cap B = \varnothing$,从而 $\vec{X}(n) \cap B = \varnothing$ 时有

$$\begin{aligned}
P\{\vec{X},\vec{X}(n+1) \cap B = \varnothing\}
&= \sum_{\vec{Y}(n)\cap B=\varnothing} P_{S_n}\{\vec{X},\vec{Y}\} P_{G_n}\{\vec{Y},G(\vec{Y}) \cap B = \varnothing\} \\
&= \sum_{\vec{Y}(n)\cap B=\varnothing} P_{S_n}\{\vec{X},\vec{Y}\}(1 - P_{G_n}\{\vec{Y},G(\vec{Y}) \cap B \neq \varnothing\}) \\
&\leqslant \sum_{\vec{Y}(n)\cap B=\varnothing} P_{S_n}\{\vec{X},\vec{Y}\}(1 - A_{G_n}) \leqslant 1 - A_{G_n}
\end{aligned}$$

则证

$$\beta_n^B \leqslant 1 - A_{G_n} \tag{3.3.8}$$

对于 $\vec{X} \cap B \neq \varnothing$,由 $\vec{Y} = S(\vec{X}) \subset \vec{X}$,从而对于 $\vec{X}(n) \cap B \neq \varnothing$ 有

$$\begin{aligned}
P\{\vec{X},\vec{X}(n+1) \cap B = \varnothing\}
&= \sum_{\vec{Y}} P_{S_n}\{\vec{X},\vec{Y}\} P_{G_n}\{\vec{Y},G(\vec{Y}) \cap B = \varnothing\} \\
&\leqslant \sum_{|\vec{Y}(X)\cap B|\geqslant m+1} P_{S_n}\{\vec{X},\vec{Y}\} P_{G_n}\{\vec{Y},G(\vec{Y}) \cap B = \varnothing\} \\
&\quad + \sum_{|\vec{Y}(X)\cap B|\geqslant m+1} P_{S_n}\{\vec{X},\vec{Y}\}[1 - P_{G_n}\{\vec{Y},G(\vec{Y}) \cap B = \varnothing\}] \\
&\leqslant 1 - \sum_{|\vec{Y}(X)|\geqslant m+1} P_{S_n}\{\vec{X},\vec{Y}\}[1 - S_{G_n}^{m+1}] \\
&\leqslant 1 - (1 - S_{G_n}^{m+1}) P_{S_n}(\vec{X},|\vec{Y}(\vec{X})| \geqslant m+1\} \\
&\leqslant 1 - (1 - S_{G_n}^{m+1}) I_{S_n}
\end{aligned} \tag{3.3.9}$$

最后一式应用定理条件(3),于是

$$\alpha_n^B \leqslant 1 - I_{S_n}(1 - S_{G_n}^{m+1}) \tag{3.3.10}$$

由式(3.3.8) 和式(3.3.10) 以及定理条件即得,对于任意满意集 $B$ 有

$$\sum_{n=1}^{\infty}(1 - \beta_n^B) \geqslant \sum_{n=1}^{\infty} A_{G_n} = \infty$$

$$0 \leqslant \frac{\alpha_n^\beta}{1 - \beta_n^\beta} \leqslant \frac{1 - I_{S_n}(1 - S_{G_n}^{m+1})}{A_{G_n}} \to 0$$

据定理 3.2.1 则证。

**定理 3.3.2** 假定遗传算法由时变比例选择算子族 $\{S_n\}$ 及组合遗传算子族 $\{C_n, m_n\}$ 构成,$S_n$ 相关的适应值调整函数为 $\sigma_n$,而 $m_n$ 所对应的变异概率为 $p_n$,若对于

$$\rho_n = \max\left\{\frac{\sigma_n(f(X))}{\sigma_n(f(Y))}; \sigma_n(f(X)) < \sigma_n(f(Y))\right\}$$

满足以下条件:

(1) $\rho_n = o\left(\frac{1}{n}\right)$

(2) $p_n = O\left(\sqrt[l]{\frac{1}{n}}\right)$

(3) $N > l$

则对于该遗传算法概率弱收敛到全局最优解集,即

$$\lim_{n\to\infty} P\{\vec{X}(n) \cap M \neq \varnothing\} = 1$$

**证明** 由例 3.3.3 知 $P_{s_n} = N - 1 = m$,且

$$I_{S_n} = (1 + (N-1)\rho_n)^{-N} = 1 - O(\rho_n) = 1 - o\left(\frac{1}{n}\right)$$

又由例 3.3.10 知

$$A_{G_n} = 1 - (1 - p_n^l)^N = 1 - \left(1 - O\left(\frac{1}{n}\right)\right)^N = O\left(\frac{N}{n}\right)$$

$$S_{G_n}^N = [1 - (1 - p_n)^l]^N = \left[1 - \left(1 - O\left(\sqrt[l]{\frac{1}{n}}\right)\right)^l\right]^N$$

$$= O(l^N n^{-\frac{N}{l}})$$

由于 $N > l$，$\left(1 - \frac{N}{l}\right) < 0$，于是

$$\frac{1 - I_{S_n}(1 - S_{G_n}^N)}{A_{G_n}} = \frac{1 - I_{S_n}}{A_{G_n}} + I_{S_n}\frac{S_{G_n}^N}{A_{G_n}}$$

$$= \frac{o\left(\frac{1}{n}\right)}{O\left(\frac{N}{n}\right)} + \left(1 - o\left(\frac{1}{n}\right)\right)\frac{O(l^N n^{-\frac{N}{l}})}{O\left(\frac{N}{n}\right)}$$

$$= o(1) + O\left(\frac{1}{N} l^N n^{1-\frac{N}{l}}\right) \to 0$$

又因为

$$\sum_{n=1}^{\infty} A_{G_n} = \sum_{n=1}^{\infty} O\left(\frac{N}{n}\right) = \infty$$

满足定理 3.3.1，则证。

**推论 3.3.1**　如果在定理 3.3.2 中取 $\sigma_n(x) = e^{-x/T_n}$，且满足

(1) $T_n = o\left(\frac{1}{\ln(n)}\right)$

(2) $p_n = O\left(\sqrt[l]{\frac{1}{n}}\right)$

(3) $N > l$

则遗传算法 $\{\vec{X}(n)\}$ 概率弱收敛到全局最优解集，即

$$\lim_{n\to\infty} P\{\vec{X}(n) \cap M \neq \varnothing\} = 1$$

**证明**　引用定理 3.3.2 中记号易证

$$\rho_n = e^{-\Delta/T_n}$$

其中

$$\Delta = \min\{f(a) - f(b);\ f(a) > f(b),\ a, b \in S\} > 0$$

$T_n = o\left(\frac{1}{\ln(n)}\right)$ 等价于 $\rho_n = o\left(\frac{1}{n}\right)$。则证。

**定理 3.3.3** 假定遗传算法由非线性 ranking 选择族 $S_n$(相应参数为 $q_n$)和组合遗传算子族$\{(C_n, m_n)\}$(变异概率为 $p_n$)构成,且满足以下条件:

(1) $q_n = 1 - o\left(\frac{1}{n}\right)$

(2) $p_n = O\left(\sqrt[l]{\frac{1}{n}}\right)$

(3) $N > l$

则该遗传算法$\{\vec{X}(n)\}$概率弱收敛到全局最优解集。

**证明** 类似定理 3.3.2。

**定理 3.3.4** 假定遗传算法由 tournament 选择算子 $S_n$ 和变异算子 $m_n$(变异概率 $p_n$)构成,且满足以下条件:

(1) $p_n = O\left(\sqrt[l]{\frac{1}{n}}\right)$

(2) tournament 尺寸 $s \geqslant l$

则遗传算法$\{\vec{X}(n)\}$概率弱收敛到全局最优解集。

**证明** 类似定理 3.3.2。

与定理3.3.1平行,可以研究遗传算法$\{\vec{X}(n)\}$概率强收敛到全局最优解集的条件。这时,需要研究对于任意满意解集

$$P_0(n) = \{\vec{X}(n) \cap B^c \neq \varnothing\} \to 0$$

为此需要利用定理 3.2.2,但这将引起一系列概念的变化。这里,我们利用鞅收敛技术研究强收敛性。

# 3-4 遗传算法的几乎处处收敛定理

前面我们讨论了遗传算法的概率收敛定理。下面我们进一步讨论遗传算法的几乎处处收敛定理。为了利用鞅收敛讨论遗传算法的几乎处处收敛定理,先给出条件期望与鞅(下鞅、上鞅)的定

义及性质。

设 $E$ 是一个基本集合，$\mathscr{F}$ 是 $E$ 上的部分子集构成的 $\sigma$ 代数，即满足：

(1) $\varnothing \in \mathscr{F}$

(2) $A \in \mathscr{F}$ 时 $A^c \in \mathscr{F}$

(3) $A_k \in \mathscr{F}(k \leqslant n)$，则 $\bigcup_{k=1}^{n} A_k \in \mathscr{F}$

(4) $A_n \in \mathscr{F}, A = \lim\limits_{n\to\infty} A_k \in \mathscr{F}$

$P$ 是 $\mathscr{F}$ 上的概率测度，则 $(E,\mathscr{F},P)$ 构成概率空间，$(E,\mathscr{F})$ 称为可测空间。若 $M$ 是可测空间 $(E_1,\mathscr{F}_1)$ 到可测空间 $(E_2,\mathscr{F}_2)$ 的映射，且对于任意 $A \in \mathscr{F}_2$ 有

$$M^{-1}(A) = \{e; M(e) \in A\} \in \mathscr{F}_1$$

则称 $M$ 为可测映射。若 $M$ 是实值可测映射，称为随机变量。

设 $(E,\mathscr{F},P)$ 是一个概率空间，$\mathscr{G} \subset \mathscr{F}$ 也是一个 $\sigma$ 代数，$f$ 是 $E$ 上的可测与可积的随机变量，$f$ 关于 $\mathscr{G}$ 的条件期望 $E(f/\mathscr{G})$ 是 $\mathscr{G}$ 可测的和可积的，且对于任意 $A \in \mathscr{G}$ 有

$$\int_A E(f/\mathscr{G})\mathrm{d}P = \int_A f\mathrm{d}P$$

显然条件期望具有以下性质：

(1) $E(E(f/\mathscr{G})) = E(f)$ (a.s.)

(2) 若 $\alpha,\beta$ 是常数，则

$$E(\alpha f + \beta g/\mathscr{G}) = \alpha E(f/\mathscr{G}) + \beta E(f/\mathscr{G}) \text{ (a.s.)}$$

若 $\mathscr{F}_n$ 是单调增的 $\sigma$ 代数列，$f_n$ 是 $E$ 上的实值函数序列，$(f_n,\mathscr{F}_n)$ 称为鞅，若 $f_n$ 关于 $\mathscr{F}_n$ 可测且可积，同时对于任意 $m < n$ 有

$$E(f_n/\mathscr{F}_m) = f_m \text{(a.s.)}$$

若 $E(f_n/\mathscr{F}_m) \geqslant f_m$(a.s.)，称为下鞅；若 $E(f_n/\mathscr{F}_m) \leqslant f_m$(a.s.)，称为上鞅。若 $\{f_n\}$ 是有界的，则对于鞅、下鞅和上鞅，$\lim\limits_{n\to\infty} f_n$ 几乎处处存在且是有界的。

现在我们利用下鞅收敛性讨论遗传算法序列。

引进以下记号:

$$b_n = \max\{P_m^n(X,Y); X \neq Y\}$$

$$\bar{a}_n = \min\{P_c^n(X \otimes Y, X); X, Y \in S\}$$

$$\bar{f}(\vec{X}) = \frac{1}{N}\sum_{i=1}^{N} f(X_i)$$

$$\mathscr{F}_n = \mathscr{F}(\vec{X}_1, \vec{X}_2, \cdots, \vec{X}_n)$$

$\mathscr{F}_n$ 表示使 $\vec{X}_1, \cdots, \vec{X}_n$ 可测的最小 $\sigma$ 代数。于是可以得到下面定理。

**定理 3.4.1** 对于例 3.2.4 中最佳值选择遗传算法,若

$$\sum_{n=1}^{\infty}[(1-\bar{a}_n) + b_n^N] < \infty \tag{3.4.1}$$

则存在单调增有界的正数序列 $\{\alpha_n\}$ 使得 $\{\bar{f}(\vec{X}(n)) + \alpha_n, \mathscr{F}_n\}$ 为非负有界下鞅,即

$$E(\bar{f}(\vec{X}(n+1)) + \alpha_{n+1}/\mathscr{F}_n) \geqslant \bar{f}(\vec{X}(n)) + \alpha_n (n \geqslant 1) \tag{3.4.2}$$

**证明** 由于 $\{\vec{X}(n)\}$ 为马尔可夫链,为证式(3.4.2) 只须证明对于任意 $\vec{X} \in S^N$,有

$$E(\bar{f}(\vec{X}(n+1)) + \alpha_{n+1}/\vec{X}(n) = \vec{X}) \geqslant \bar{f}(\vec{X}) + \alpha_n (n \geqslant 1) \tag{3.4.3}$$

首先,由式(3.2.7) 有

$$\begin{aligned} & E(\bar{f}(\vec{X}(n+1))/\vec{X}(n) = \vec{X}) \\ & = \sum_{\vec{Z}} \bar{f}(\vec{Z}) P\{\vec{X}(n+1) = \vec{Z}/\vec{X}(n) = \vec{X}\} \\ & = \frac{1}{N}\sum_{j=1}^{N}\sum_{\vec{Z}} f(Z_j) P\{\vec{X}(n+1) = \vec{Z}/\vec{X}(n) = \vec{X}\} \\ & = \sum_{Z_1} f(Z_1) P\{\vec{X}(n+1)_1 = Z_1/\vec{X}(n) = \vec{X}\} \\ & = \sum_{\vec{Y}} P_m^n\{\vec{X}, \vec{Y}\} \sum_{X, Y \in B(\vec{Y})} (|\vec{Y}(X)| |\vec{Y}(Y)| / |B(\vec{Y})|^2 \end{aligned}$$

$$\cdot \sum_{Z} P_c^n \{X \otimes Y, Z\} f(Z))$$

因而

$$E(\bar{f}(\vec{X}(n+1))/\vec{X}(n) = \vec{X}) - \bar{f}(\vec{X})$$
$$= \sum_{\vec{Y}} P_m^n \{\vec{X}, \vec{Y}\} \sum_{X, Y \in B(\vec{Y})} \frac{|\vec{Y}(X)| \, |\vec{Y}(Y)|}{|B(\vec{Y})|^2}$$
$$\cdot \sum_{Z} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X})) \qquad (3.4.4)$$

为了估计式(3.4.4),分以下几种情况讨论:

(1) 若 $\vec{X} \notin \beta^*$,其中 $\beta^*$ 表示适应值全等的种群全体,则

$$\hat{f}(\vec{X}) - \bar{f}(\vec{X}) \geqslant \frac{\delta}{N}$$

其中

$$\delta = \min\{|f(X) - f(Y)|;\ f(X) \neq f(Y)\}$$

记

$$\Sigma_1 = \sum_{\vec{Y} \neq \vec{X}} P_m^n \{\vec{X}, \vec{Y}\} \sum_{X, Y \in B(Y)} P_s \{\vec{Y}, X \otimes Y\}$$
$$\cdot \sum_{Z} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$
$$\Sigma_2 = P_m^n \{\vec{X}, \vec{X}\} \sum_{X, Y \in B(\vec{X})} P_s \{\vec{X}, X \otimes Y\}$$
$$\cdot \sum_{Z} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$

则

$$E(\bar{f}(\vec{X}(n+1))/\vec{X}(n) = \vec{X}) - \bar{f}(\vec{X}) = \Sigma_1 + \Sigma_2$$

记

$$\Delta = \max\{|f(X) - f(Y)|;\ X, Y \in S\}$$

则对于任意 $Z \in S$ 有 $f(Z) - \bar{f}(\vec{X}) \geqslant -\Delta$,于是有

$$\Sigma_1 \geqslant -\Delta \sum_{\vec{Y} \neq \vec{X}} P_m^n (\vec{X}, \vec{Y})$$
$$= -\Delta(1 - P_m^n \{\vec{X}, \vec{X}\})$$

$$\geqslant -\Delta(1-a_n^N) \tag{3.4.5}$$

其中

$$a_n = \min\{P_m^n\{X,X\};X\in S\}$$

$$\Sigma_2 = P_m^n\{\vec{X},\vec{X}\}\sum_{X,Y\in B(\dot{X})}P_s(\vec{X},X\otimes Y)\cdot\sum_{Z\in B(\dot{X})}P_c^n\{X\otimes Y,Z\}(f(Z)-\bar{f}(\vec{X}))$$
$$+P_m^n\{\vec{X},\vec{X}\}\sum_{X,Y\in B(\dot{X})}P_s\{\vec{X},X\otimes Y\}\cdot\sum_{Z\notin B(\dot{X})}P_c^n\{X\otimes Y,Z\}(f(Z)-\bar{f}(\vec{X}))$$

若 $Z\in B(\vec{X})$,则 $f(Z)-\bar{f}(\vec{X})=\hat{f}(\vec{X})-\bar{f}(\vec{X})\geqslant\frac{\delta}{N}$,于是

$$\Sigma_2\geqslant a_n^N\frac{\delta}{N}\sum_{X,Y\in B(\dot{X})}P_s\{\vec{X},X\otimes Y\}\sum_{Z\in B(\dot{X})}P_c^n\{X\otimes Y,Z\}$$
$$-\Delta\sum_{X,Y\in B(\dot{X})}P_s\{\vec{X},X\otimes Y\}\sum_{Z\notin B(\dot{X})}P_c^n\{X\otimes Y,Z\}$$
$$\geqslant a_n^N\frac{\delta}{N}\sum_{X,Y\in B(\dot{X})}P_s\{\vec{X},X\otimes Y\}P_c^n\{X\otimes Y,X\}$$
$$-\Delta\sum_{X,Y\in B(\dot{X})}P_s\{\vec{X},X\otimes Y\}(1-P_c^n\{X\otimes Y,X)\}$$
$$\geqslant a_n^N\frac{\delta}{N}\bar{a}_n-\Delta(1-\bar{a}_n)=c_1-\Delta(1-\bar{a}_n) \tag{3.4.6}$$

由式(3.4.5)和(3.4.6)即得

$$E(\bar{f}(\vec{X}(n+1))/\vec{X}(n)=\vec{X})-\bar{f}(\vec{X})$$
$$>c_1-\Delta(1-a_n^N)-\Delta(1-\bar{a}_n)$$

由定理条件知,$a_n\to 1,\bar{a}_n\to 1$,于是不妨假定

$$E(\bar{f}(\vec{X}(n+1))/\vec{X}(n)=\vec{X})-\bar{f}(\vec{X})>c_1$$

(2) 若 $\vec{X}\in\beta^*\setminus M$,则 $f(X_1)=\cdots=f(X_N)=\bar{f}(\vec{X})=\hat{f}(\vec{X})<f^*$,记

$$\Sigma_1 = P_m^n\{\vec{X},\vec{X}\}\sum_{X,Y\in B(\vec{X})}P_s\{\vec{X},X\otimes Y\}$$

$$\cdot \sum_{Z} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$

$$\Sigma_2 = \sum_{\substack{\dot{Y} \neq \dot{X} \\ \dot{Y} \cap \dot{X} \neq \varnothing}} P_m^n \{\vec{X}, \vec{Y}\} \sum_{X, Y \in B(\dot{X})} P_s \{\vec{Y}, X \otimes Y\}$$

$$\cdot \sum_{Z} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$

$$\Sigma_3 = \sum_{\dot{Y} \cap \dot{X} = \varnothing} P_m^n \{\vec{X}, \vec{Y}\} \sum_{X, Y \in B(\dot{Y})} P_s \{\vec{Y}, X \otimes Y\}$$

$$\cdot \sum_{Z} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$

则

$$E(\bar{f}(\vec{X}(n+1))/\vec{X}(n) = \vec{X}) - \bar{f}(\vec{X})$$

$$= \Sigma_1 + \Sigma_2 + \Sigma_3$$

若 $X \in B(\vec{X})$，则 $f(X) = \bar{f}(\vec{X}) = \hat{f}(\vec{X})$，则

$$\Sigma_1 = P_m^n \{\vec{X}, \vec{X}\} \sum_{X, Y \in B(\dot{X})} P_s \{\vec{X}, X \otimes Y\}$$

$$\cdot \sum_{Z \neq X} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$

$$\geqslant -\Delta P_m^n \{\vec{X}, \vec{X}\} \sum_{X, Y \in B(\vec{X})} P_s \{\vec{X}, X \otimes Y\} \sum_{Z \neq X} P_c^n \{X \otimes Y, Z\}$$

$$\geqslant -\Delta P_m^n \{\vec{X}, \vec{X}\} \sum_{X, Y \in B(\vec{X})} P_s \{\vec{X}, X \otimes Y\}$$

$$\cdot (1 - P_c^n \{X \otimes Y, X\})$$

$$\geqslant -\Delta(1 - \bar{a}_n) = -c_2(1 - \bar{a}_n) \tag{3.4.7}$$

对于

$$\vec{Y}' = (X_1, \cdots, X_{N-1}, e)$$

其中 $e \neq X_N$，$f(e) = f^*$，$B(\vec{Y}') = \{e\}$，则有

$$P_m^n \{\vec{X}, \vec{Y}'\} \sum_{X, Y \in B(\dot{X})} P_s \{\vec{Y}', X \otimes Y\}$$

$$\cdot \sum_{Z} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$

$$= P_m^n \{\vec{X}, \vec{Y}'\} P_s \{\vec{Y}', e \times e\} P_c^n \{e \times e, e\}$$

$$\cdot (f(e) - \bar{f}(\vec{X}))$$
$$\geqslant P_m^n\{\vec{X}, \vec{Y}'\}(\hat{f}(\vec{X}) - \bar{f}(\vec{X}))$$
$$\geqslant a_n^{N-1} d_n \delta = c_3 d_n$$

其中

$$d_n = \min\{P_m^n\{X, Y\}; X \neq Y\}$$

另一方面，若 $\vec{Y} \cap \vec{X} \neq \varnothing$，对于 $Y \in B(\vec{Y})$，$f(Y) = \hat{f}(\vec{Y}) \geqslant \hat{f}(\vec{X}) = \bar{f}(\vec{X})$，从 $f(Z) < \bar{f}(\vec{X})$ 得到 $Z \neq Y$，这样

$$\Sigma_2 \geqslant c_3 d_n + \sum_{\substack{\vec{Y} \cap \vec{X} \neq \varnothing \\ Y' \neq Y \neq X}} P_m^n\{\vec{X}, \vec{Y}\} \sum_{X, Y \in B(\vec{Y})} P_s\{\vec{Y}, X \otimes Y\}$$
$$\cdot \sum_{Z \neq X} P_c^n\{X \otimes Y, Z\}(f(Z) - \bar{f}(\vec{X}))$$
$$\geqslant c_3 d_n - \Delta \sum_{\substack{\vec{Y} \cap \vec{X} \neq \varnothing \\ Y \neq X}} P_m^n\{\vec{X}, \vec{Y}\}$$
$$\cdot \sum_{X, Y \in B(\vec{Y})} P_s\{\vec{Y}, X \otimes Y\}(1 - P_c^n(X \otimes Y, X))$$
$$\geqslant c_3 d_n - \Delta(1 - \bar{a}_n) \sum_{\substack{\vec{Y} \cap \vec{X} \neq \varnothing \\ Y \neq X}} P_m^n\{\vec{X}, \vec{Y}\}$$
$$\geqslant c_3 d_n - \Delta(1 - \bar{a}_n)(1 - P_m^n(\vec{X}, \vec{X}))$$
$$= c_3 d_n - \Delta(1 - \bar{a}_n)(1 - a_n^N)$$
$$= c_3 d_n - \Delta(1 - \bar{a}_n) \tag{3.4.8}$$

当 $\vec{X} \cap \vec{Y} = \varnothing$ 时，$X_i \neq Y_i (i \leqslant N)$，于是

$$\Sigma_3 \geqslant -\Delta \sum_{\vec{Y} \cap \vec{X} = \varnothing} P_m^n\{\vec{X}, \vec{Y}\}$$
$$\geqslant -\Delta |S|^N b_n^N = -c_4 b_n^N \tag{3.4.9}$$

联合式(3.4.7)，(3.4.8)，(3.4.9) 即得

$$E(\bar{f}(\vec{X}(n+1))/\vec{X}(n) = \bar{X}) - \bar{f}(\vec{X})$$
$$\geqslant c_3 d_n - c_2(1 - \bar{a}_n) - c_4 b_n^N \tag{3.4.10}$$

(3) 若 $\vec{X} \subset M$，则 $\bar{f}(\vec{X}) = \hat{f}(\vec{X}) = f^*$，记

$$\Sigma_1 = \sum_{\dot{Y} \cap \dot{X} = \varnothing} P_m^n \{\vec{X}, \vec{Y}\} \sum_{X, Y \in B(\dot{Y})} P_s \{\vec{Y}, X \otimes Y\} \cdot \sum_{Z} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$

$$\Sigma_2 = \sum_{\dot{Y} \cap \dot{X} \neq \varnothing} P_m^n \{\vec{X}, \vec{Y}\} \sum_{X, Y \in B(\dot{Y})} P_s \{\vec{Y}, X \otimes Y\} \cdot \sum_{Z \in B(\dot{Y})} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$

$$\Sigma_3 = \sum_{\dot{Y} \cap \dot{X} \neq \varnothing} P_m^n \{\vec{X}, \vec{Y}\} \sum_{X, Y \in B(\dot{Y})} P_s \{\vec{Y}, X \otimes Y\} \cdot \sum_{Z \notin B(\dot{Y})} P_c^n \{X \otimes Y, Z\} (f(Z) - \bar{f}(\vec{X}))$$

当$\vec{X} \cap \vec{Y} \neq \varnothing, \bar{f}(\vec{Y}) = \bar{f}(\vec{X}) = f^*$时，若$Z \in B(\vec{Y})$，则$f(Z) - \bar{f}(\vec{X}) = 0$，于是$\Sigma_2 = 0$。当$X \in B(\vec{Y}), Z \notin B(\vec{Y})$时$Z \neq X$，于是

$$\Sigma_3 \geqslant \sum_{\dot{Y} \cap \dot{X} \neq \varnothing} P_m^n \{\vec{X}, \vec{X}\} \cdot (-\Delta(1 - P_c^n \{X \otimes Y, X\}))$$
$$\geqslant -\Delta(1 - \bar{a}_n) \sum_{\dot{Y} \cap \dot{X} \neq \varnothing} P_m^n \{\vec{X}, \vec{Y}\}$$

于是

$$E(\bar{f}(\vec{X}(n+1)) / \vec{X}(n) = \vec{X}) - \bar{f}(\vec{X})$$
$$\geqslant -\Delta \sum_{\dot{X} \cap \dot{Y} = \varnothing} P_m^n \{\vec{X}, \vec{Y}\} - \Delta(1 - \bar{a}_n) \cdot \sum_{\dot{X} \cap \dot{Y} \neq \varnothing} P_m^n \{\vec{X}, \vec{Y}\}$$
$$\geqslant -\Delta |S|^N b_n^N - \Delta(1 - \bar{a}_n)$$
$$= -c_5(b_n^N + (1 - \bar{a}_n)) \tag{3.4.11}$$

令$c = (c_2 + c_4 + c_5)$，综合以上三种情形，有

$$E(\bar{f}(\vec{X}(n+1)) / \vec{X}(n) = \vec{X}) - \bar{f}(\vec{X})$$
$$\geqslant -c_2(1 - \bar{a}_n) - c_4 b_n^N - c_5(b_n^N + 1 - \bar{a}_n)$$
$$\geqslant -c(b_n^N + 1 - \bar{a}_n)$$

令

$$\alpha_n = c\sum_{k=1}^{n-1}(1-\bar{a}_k+b_k^N)$$

$\alpha_n$ 是正的单调增的有界数列，并满足

$$E(\bar{f}(\vec{X}(n+1))+\alpha_{n+1}/\vec{X}(n)=X)\geqslant\bar{f}(\vec{X})+\alpha_n$$

于是定理 3.4.1 得证。

**定理 3.4.2** 对于例 3.2.5 中最优值选择遗传算法，若满足

(1) $\sum_{n=1}^{\infty}(1-\bar{a}_n+b_n^N)<\infty$

(2) $\sum_{n=1}^{\infty}d_n=\infty$

则 $\bar{f}(\vec{X}(n))$ 几乎处处收敛到全局最优解，即

$$P\{\lim_{n\to\infty}\bar{f}(\vec{X}(n))=f^*\}=1$$

同时有

$$P\{\lim_{n\to\infty}[\vec{X}(n)\subset M]\}=1$$

即 $\{\vec{X}(n)\}$ 几乎处处强收敛到全局最优解集。

**证明** 由定理 3.4.1 知 $\bar{f}(\vec{X}(n))+\alpha_n$ 是下鞅，且 $\alpha_n$ 存在有限极限，于是 $\bar{f}(\vec{X}(n))$ 几乎处处收敛，记

$$\lim_{n\to\infty}\bar{f}(\vec{X}(n))=\xi \quad (\text{a.e.})$$

利用条件期望的性质，有

$$\begin{aligned}
&E(\bar{f}(\vec{X}(n+1)))-E(\bar{f}(\vec{X}(n)))\\
&\quad=E[E(\bar{f}(\vec{X}(n+1))/\vec{X}(n))]-E[\bar{f}(\vec{X}(n))]\\
&\quad=\sum_{\vec{X}}P\{\vec{X}(n)=\vec{X}\}\\
&\qquad\cdot(E(\bar{f}(\vec{X}(n+1))/\vec{X}(n)=\vec{X}]-\bar{f}(\vec{X}))\\
&\quad=\sum_{\vec{X}\in\beta^*}P\{\vec{X}(n)=\vec{X}\}\\
&\qquad\cdot(E[\bar{f}(\vec{X}(n+1))/\vec{X}(n)=\vec{X}]-\bar{f}(\vec{X}))\\
&\qquad+\sum_{\vec{X}\in\beta^*\setminus M}P\{\vec{X}(n)=\vec{X}\}
\end{aligned}$$

$$\cdot(E[\bar{f}(\vec{X}(n+1))/\vec{X}(n)=\vec{X}]-\bar{f}(\vec{X}))$$

$$+\sum_{\vec{X}\in M}P\{\vec{X}(n)=\vec{X}\}$$

$$\cdot(E[\bar{f}(\vec{X}(n+1))/\vec{X}(n)]-\bar{f}(\vec{X}))$$

$$\geqslant c_1P\{\vec{X}(n)\notin\beta^*\}+(c_3d_n-c_2(1-\bar{a}_n)-c_4b_n^N)$$

$$\cdot P\{\vec{X}(n)\in\beta^*\setminus M\}$$

$$-(c_5b_n^*+1-\bar{a}_n)P(\vec{X}(n)\in M)$$

$$\geqslant c_1P\{\vec{X}(n)\notin\beta^*\}+c_3d_nP\{\vec{X}(n)\in\beta^*\setminus M\}$$

$$-c(1-\bar{a}_n+b_n^N)P\{\vec{X}(n)\in\beta^*\}$$

其中

$$c=(c_2+c_4+c_5)$$

关于上式对 $n$ 求和,即得

$$f^*\geqslant E[\bar{f}(\vec{X}(n+1))]-E[\bar{f}(\vec{X}(1))]$$

$$\geqslant c_1\sum_{k=1}^{n}P\{\vec{X}(k)\notin\beta^*\}+c_3\sum_{k=1}^{n}d_kP\{\vec{X}(k)\in\beta^*\setminus M\}$$

$$-c\sum_{k=1}^{n}(1-\bar{a}_k+b_k^N)P\{\vec{X}(k)\in\beta^*\}$$

令 $n\to\infty$,有

$$f^*+c\sum_{n=1}^{\infty}(1-\bar{a}_n+b_n^N)P\{\vec{X}(n)\in\beta^*\}$$

$$\geqslant c_1\sum_{n=1}^{\infty}P\{\vec{X}(n)\notin\beta^*\}+c_3\sum_{n=1}^{\infty}d_nP\{\vec{X}(n)\in\beta^*\setminus M\}$$

由定理条件,有

$$\sum_{n=1}^{\infty}P\{\vec{X}(n)\notin\beta^*\}<\infty \tag{3.4.12}$$

$$\sum_{n=1}^{\infty}d_nP\{\vec{X}(n)\in\beta^*\setminus M\}<\infty \tag{3.4.13}$$

从式(3.4.12)和(3.4.13),有

$$\sum_{n=1}^{\infty} d_n P\{\vec{X}(n) \notin M\} < \infty$$

由于 $\xi \leqslant f^*$(a.s.)，只须证明 $P\{\xi < f^*\} = 0$。若 $P\{\xi < f^*\} > 0$，则存在正数 $\epsilon$ 和 $a$，使

$$P\{\xi < f^* - 2\epsilon\} > 2a$$

由于 $f(\vec{X}(n))$ 几乎处处收敛到 $\xi$，基于叶果洛夫定理，存在可测集 $A$，$P(A) > 1 - a$，$\bar{f}(\vec{X}(n))$ 在 $A$ 上是一致收敛的。这样，对于任意 $\epsilon > 0$，存在 $N_0$（不依赖于 $A$），使当 $n \geqslant N_0$ 时总有

$$\xi - \epsilon < \bar{f}(\vec{X}(n)) < \xi + \epsilon$$

令

$$B = A \cap \{\xi < f^* - 2\epsilon\}$$

则当 $P(B) \geqslant a$ 且 $n \geqslant N_1$ 时，有

$$\bar{f}(\vec{X}(n)) < \xi + \epsilon < f^* - \epsilon \Rightarrow \vec{X}(n) \notin M$$

因而当 $n \geqslant N_1$ 时，$P\{\vec{X}(n) \notin M\} \geqslant a$，

$$\sum_{n=1}^{\infty} d_n P\{\vec{X}(n) \notin M\} \geqslant a \sum_{n=1}^{\infty} d_n = \infty$$

与式(3.4.13)矛盾，则有

$$P\{\lim_{n\to\infty} \bar{f}(\vec{X}(n)) = f^*\} = 1$$

再由定理3.1.6，即有

$$P\{\lim_{n\to\infty} [\vec{X}(n) \subset M]\} = 1$$

**推论3.4.1** 在定理3.4.2中，若采用单点杂交或均匀杂交，杂交概率为 $q_n$，若

$$\sum_{n=1}^{\infty} (p_n^N + q_n) < \infty, \quad \sum_{n=1}^{\infty} p_n^l = \infty$$

则 $\{\vec{X}(n)\}$ 几乎处处强收敛到全局最优解集。

**证明** 若变异为

$$P_m^n\{X, Y\} \doteq p_n^{d(X,Y)} (1 - p)^{l-d(X,Y)}$$

则

$$a_n = (1-p_n)^l$$
$$b_n = (1-p_n)^{l-1}p_n = O(p_n)$$
$$d_n = p_n^l$$

对于单点杂交运算与均匀杂交运算均有

$$\bar{a}_n = 1 - q_n$$

则
$$\sum_{n=1}^{\infty}[1-\bar{a}_n + b_n^N] = \sum_{n=1}^{\infty}[q_n + O(p_n^N)] < \infty$$
$$\sum_{n=1}^{\infty} d_n = \sum_{n=1}^{\infty} p_n^l = \infty$$

**推论 3.4.2** 对于仅有变异和选择的遗传算法，当 $\sum_{n=1}^{\infty} b_n^N < \infty$ 时定理 3.4.1 成立，若更有 $\sum_{n=1}^{\infty} d_n = \infty$，则定理 3.4.2 成立。

**证明** 直接利用定理即证。

对于例 3.2.3 采用父子两代竞争选择方式，我们同样可以用鞅讨论遗传算法几乎处处收敛性。

**定理 3.4.3** 设 $\{\vec{X}(n)\}$ 是例 3.2.3 中的父代参与竞争无杂交遗传算法，则 $(\bar{f}(\vec{X}(n)),\mathscr{F}_n)$ 是正的有界下鞅，即

$$E(\bar{f}(\vec{X}(n+1))/\mathscr{F}_n) \geqslant \bar{f}(\vec{X}(n)) \quad (n \geqslant 1)$$

**证明** 由于 $\vec{Z} \in \langle \vec{X} \oplus \vec{Y} \rangle$ 时，$\bar{f}(\vec{Z}) \geqslant \bar{f}(\vec{X})$，则

$$
\begin{aligned}
&E(\bar{f}(\vec{X}(n+1))/\vec{X}(n) = \vec{X}) \\
&= \sum_{\vec{Z}} \bar{f}(\vec{Z}) P\{\vec{X}(n+1) = \vec{Z}/\vec{X}(n) = \vec{X}\} \\
&= \sum_{Z} \bar{f}(Z) \sum_{Y} P_m^n\{\vec{X},\vec{Y}\} P_s\{\langle \vec{X} \oplus \vec{Y} \rangle, \vec{Z}\} \\
&= \sum_{Y} P_m^n\{\vec{X},\vec{Y}\} \sum_{\vec{Z} \in \langle \vec{X} \oplus \vec{Y} \rangle} \bar{f}(Z) \\
&\quad \cdot \prod_{k=1}^{N} \left[ f(Z_k) / \sum_{\vec{Y} \in \langle \vec{X} + \vec{Y} \rangle} f(Y) \right]
\end{aligned}
$$

$$\geqslant \bar{f}(\vec{X})$$

则证。

**定理 3.4.4** 在定理 3.4.3 条件下,记

$$r_n^* = \min_{\vec{X},\vec{Y}} P_m^n\{\vec{X},\vec{Y}\} \quad (n \geqslant 1)$$

当 $\sum_{n=1}^{\infty} r_n^* = \infty$ 时,有

(1) $\lim_{n\to\infty} \bar{f}(\vec{X}(n)) = f^*$ (a.s.)

(2) $P\{\lim_{n\to\infty}[\vec{X}(n) \subset M]\} = 1$

**证明** 令

$$F_n = \{\vec{X}(n) \cap M^c \neq \varnothing\} = \{\bar{f}(\vec{X}(n)) \neq f^*\}$$

对于下鞅 $\bar{f}(\vec{X}(n))$,利用条件期望性质有

$$f^* P\{F_n^c\} \geqslant \int_{F_n^c} \bar{f}(\vec{X}(n+1))\mathrm{d}P \geqslant \int_{F_n^c} \bar{f}(\vec{X}(n))\mathrm{d}P$$

$$= f^* P\{F_n^c\}$$

于是

$$\int_{F_n^c} \bar{f}(\vec{X}(n+1))\mathrm{d}P = f^* P\{F_n^c\}$$

由于 $0 \leqslant \bar{f}(\vec{X}) \leqslant f^*$,在 $F_n^c$ 上 $\bar{f}(\vec{X}(n+1)) = f^*$ (a.s.)。于是 $F_n^c \subset F_{n+1}^c$,也即 $F_{n+1} \subset F_n$。另外,若 $\vec{X} \cap M^c \neq \varnothing$,于是存在 $X_k$ 使 $f(X_k) < f^*$。令 $\vec{Y}$ 使 $\vec{Y} \cap M^c = \varnothing$,记此相应的 $\vec{Y}$ 为 $\vec{Y}_X$,则

$$\bar{f}(\langle \vec{X} \oplus \vec{Y}_X \rangle) \geqslant \bar{f}(\vec{X}) + \frac{1}{N}\delta$$

于是,由条件期望性质即得

$$E(\bar{f}(\vec{X}(n+1))) - E(\bar{f}(\vec{X}(n)))$$

$$= E(E(\bar{f}(\vec{X}(n+1))/\vec{X}(n))) - E(\bar{f}(\vec{X}(n)))$$

$$= \sum_{\vec{X}} P\{\vec{X}(n) = \vec{X}\}$$

$$\cdot \sum_{\vec{Y}} P_m^n\{\vec{X},\vec{Y}\}(\bar{f}(\langle\vec{X}\oplus\vec{Y}\rangle)-\bar{f}(\vec{X}))$$

$$\geqslant \sum_{\vec{X}\cap M\neq\varnothing} P\{\vec{X}(n)=\vec{X}\}P_m^n\{\vec{X},\vec{Y}_X\}$$

$$\cdot(\bar{f}(\langle\vec{X}\oplus\vec{Y}\rangle)-\bar{f}(\vec{X}))$$

$$\geqslant \frac{\delta}{N}P\{F_n\}r_n^*$$

对上面不等式求和，有

$$f^*\geqslant E(\bar{f}(\vec{X}(n+1)))-E(\bar{f}(\vec{X}(1)))$$

$$\geqslant \frac{\delta}{N}\sum_{k=1}^{n}P\{F_k\}r_k^*$$

令 $n\to\infty$，即得

$$\sum_{n=1}^{\infty}P\{F_n\}r_n^*<\infty$$

由于 $\sum_{n=1}^{\infty}r_n^*=\infty$，于是必有 $P\{F_n\}\to 0$，即

$$\lim_{n\to\infty}\bar{f}(\vec{X}(n))=f^*\ (\text{a.s.})$$

则证(1)。

(2) 由定理3.1.6可证。

我们指出，当变异转移概率为

$$P_m^n\{\vec{X},\vec{Y}\}=\prod_{j=1}^{N}P_m^n\{X_j,Y_j\}$$

时 $r_n^*=r_n^N$，于是定理3.4.2的条件与定理3.2.5的条件一致。事实上，定理3.2.5在 $\sum_{n=1}^{\infty}r_n^*=\infty$ 条件下同样成立。

对于标准遗传算法，变异概率 $p_n\equiv p>0$ 时，$r_n^*=p^{lN}(p<1/2)$，于是定理3.4.4条件满足。因此，对于变异不变遗传算法采用定理3.4.4的算法几乎处处强收敛到全局最优解集。

## 3-5 遗传算法的渐近收敛定理

遗传算法的收敛定理是对一串遗传机制序列 $G_n$ 进行讨论，这时 $\vec{X}(n+1)=G_n(\vec{X}(n))$。遗传算法的渐近收敛定理是对遗传机制序列 $G_n^k$ 进行讨论，这时

$$\vec{X}^k(n+1)=G_n^k(\vec{X}^k(n))$$

**定义 3.5.1** 称遗传算法的渐近序列概率弱收敛到全局最优解集，若

$$\lim_{k\to\infty}\lim_{n\to\infty}P\{(\vec{X}^k(n))\cap M\neq\varnothing\}=1$$

称遗传算法的渐近序列概率强收敛到全局最优解集，若

$$\lim_{k\to\infty}\lim_{n\to\infty}P\{(\vec{X}^k(n))\cap M^c=\varnothing\}=1$$

称遗传算法的渐近序列是几乎处处弱收敛到全局最优解集，若

$$P\{\lim_{k\to\infty}\lim_{n\to\infty}[(\vec{X}^k(n))\cap M\neq\varnothing]\}=1$$

称遗传算法的渐近序列是几乎处处强收敛到全局最优解集，若

$$P\{\lim_{k\to\infty}\lim_{n\to\infty}[(\vec{X}^k(n))\cap M^c=\varnothing]\}=1$$

**定理 3.5.1** 对于遗传算法的渐近序列 $\{\vec{X}^k(n)\}$，若满足

$$P\{\vec{X}^k(n+1)\cap M=\varnothing/\vec{X}^k(n)=\vec{X}\}$$
$$\leqslant\begin{cases}\alpha_k, & \vec{X}\cap M\neq\varnothing\\ \beta_k, & \vec{X}\cap M=\varnothing\end{cases}$$

且 $\beta_k<1(k\geqslant1)$，$\dfrac{\alpha_k}{1-\beta_k}\to0(k\to\infty)$，则 $\{\vec{X}^k(n)\}$ 渐近概率弱收敛到最优解集。若对于 $\{\vec{X}^k(n)\}$ 满足

$$P\{\vec{X}^k(n+1)\cap M^c\neq\varnothing/\vec{X}^k(n)=\vec{X}\}$$
$$\leqslant\begin{cases}\alpha_k', & \vec{X}\cap M^c=\varnothing\\ \beta_k', & \vec{X}\cap M^c\neq\varnothing\end{cases}$$

且 $\beta_k < 1(k \geqslant 1)$，$\frac{\alpha_k}{1-\beta_k} \to 0(k \to \infty)$，则 $\{\vec{X}^k(n)\}$ 渐近概率强收敛到最优解集。

**证明**　设

$$P_k(n) = P\{\vec{X}^k(n) \cap M \neq \varnothing\}$$

易证　$P_k(n+1) \leqslant \alpha_k + \beta_k P_k(n)$

从而

$$P_k(n+1) \leqslant \sum_{j=0}^{n-1} \beta_k^j \alpha_k + \beta_k^n P_0(1) \leqslant \frac{\alpha_k}{1-\beta_k} + \beta_k^n$$

于是

$$\overline{\lim_{k\to\infty}}\ \overline{\lim_{n\to\infty}} P_k(n+1) \leqslant 0$$

证得

$$\lim_{k\to\infty} \lim_{n\to\infty} P\{(\vec{X}^k(n)) \cap M \neq \varnothing\} = 1$$

即 $\{\vec{X}^k(n)\}$ 渐近概率弱收敛到最优解集。同理可证 $\{\vec{X}^k(n)\}$ 渐近概率强收敛到最优解集。

下面讨论渐近模拟退火遗传算法的收敛性。

渐近模拟退火遗传算法分为两步：

(1) 通过变异 $m_k$ 使 $\vec{X}^k(n)$ 变为 $\vec{Y}^k(n)$，变异算子 $m_k$ 转移概率为

$$P_k\{X, Y\} = P\{m_k(X) = Y\} = \alpha_k^{d(X,Y)}(1-\alpha_k)^{l-d(X,Y)}$$

其中 $\alpha_k = k^{-a}$，$a$ 是正实数。变异 $m_k$ 的种群转移概率为

$$P_{m_k}\{\vec{X}, \vec{Y}\} = \prod_{j=1}^{N} P_k\{X_j, Y_j\}$$

显然有

$$\lim_{\alpha_k\to 0} P_{m_k}\{\vec{X}, \vec{Y}\} = \prod_{j=1}^{N} \delta(X_j, Y_j) \tag{3.5.1}$$

其中

$$\delta(X,Y)=\begin{cases}1, & X=Y\\ 0, & X\neq Y\end{cases}$$

由此可见，$m_k$ 是对不变异的一种扰动，它依赖于 $\alpha_k$，当 $\alpha_k\to 0$ 时扰动逐渐消失。

(2) 通过选择 $S_k$ 使 $\vec{X}^k(n)\oplus\vec{Y}^k(n)$ 变为 $\vec{X}^k(n+1)$，其中 $\vec{X}^k(n)\oplus\vec{Y}^k(n)$ 表示由 $\vec{X}^k(n)$ 与 $\vec{Y}^k(n)$ 联合而成的 $2N$ 个个体组成的向量。选择算子 $S_k$ 的转移概率为

$$P_{S_k}\{\vec{X}\oplus\vec{Y},Z_i\}=P\{S_k(\vec{X}\oplus\vec{Y})_i=Z_i\}$$

$$=\frac{(\vec{X}\oplus\vec{Y})(Z_i)\cdot J_k(f(Z_i))}{\sum_{j=1}^{N}(J_k(f(X_j))+J_k(f(Y_j)))}\quad(Z_i\in\vec{X}\oplus\vec{Y})$$

其中 $(\vec{X}\oplus\vec{Y})(Z_i)$ 表示 $Z_i$ 在 $(\vec{X}\oplus\vec{Y})$ 中出现的次数。$J_k:R^+\to R^+$ 是严格增函数。可取

$$J_k(f(Z_i))=\exp\{-f(Z_i)\ln\beta_k\}$$

$0<\beta_k<1$，如取 $\beta_k=k^{-b}$，$b$ 为正实数。于是通过选择算子种群转移概率为

$$P_{S_k}\{\vec{X}\oplus\vec{Y},\vec{Z}\}=P\{S_k(\vec{X}\oplus\vec{Y})=\vec{Z}\}$$

$$=\prod_{j=1}^{N}P_{S_k}\{\vec{X}\oplus\vec{Y},Z_i\}$$

易证

$$\lim_{\beta_k\to 0}P_{S_k}\{\vec{X}\oplus\vec{Y},\vec{Z}\}=\prod_{j=1}^{N}\frac{(\vec{X}\oplus\vec{Y})(\hat{Z}_i)}{|B(\vec{X}\oplus\vec{Y})|}\qquad(3.5.2)$$

其中 $B(\vec{X}\oplus\vec{Y})$ 表示在 $(\vec{X}\oplus\vec{Y})$ 中取最大值的个体，$(\vec{X}\oplus\vec{Y})(\hat{Z}_i)$ 表示 $Z_i$ 在 $(\vec{X}\oplus\vec{Y})$ 中的个数。

下面，我们进一步分析 $k\to\infty$ 时 $m_k$ 和 $S_k$ 的转移概率的性质。当 $k$ 充分大时，有

$$P_{m_k}\{\vec{X},\vec{Y}\}=(1-k^{-a})^{Nl-d(\vec{X},\vec{Y})}k^{-ad(\vec{X},\vec{Y})}$$

$$\approx \exp\{-ad(\vec{X},\vec{Y})\ln k\} \tag{3.5.3}$$

其中

$$d(\vec{X},\vec{Y}) = \sum_{j=1}^{N} d(X_j, Y_j)$$

$$\begin{aligned}
&P_{S_k}\{\vec{X}\oplus\vec{Y},\vec{Z}\} \\
&= \prod_{j=1}^{N} \frac{(\vec{X}\oplus\vec{Y})(Z_j)\exp\{bf(Z_j)\ln k\}}{\sum_{i=1}^{N}[\exp\{bf(X_i)\ln k\} + \exp\{bf(Y_i)\ln k\}]} \\
&\approx \prod_{j=1}^{N} \frac{(\vec{X}\oplus\vec{Y})(Z_j)\exp\{bf(Z_j) - \hat{f}(\vec{X}\oplus\vec{Y})\ln k\}}{|B(\vec{X}\oplus\vec{Y})|} \\
&= C(\vec{X},\vec{Y};\vec{Z})\exp\{b\sum_{j=1}^{N}(f(Z_j) - \hat{f}(\vec{X}\oplus\vec{Y}))\ln k\}
\end{aligned} \tag{3.5.4}$$

其中

$$C(\vec{X},\vec{Y};\vec{Z}) = \prod_{j=1}^{N} \frac{(\vec{X}\oplus\vec{Y})(Z_j)}{|B(\vec{X}\oplus\vec{Y})|}$$

联合式(3.5.3) 与(3.5.4) 即得 $k\to\infty$ 时有

$$\begin{aligned}
&P\{\vec{X}^k(n+1) = \vec{Z}/\vec{X}^k(n) = \vec{X}\} \\
&= \sum_{\vec{Y}} P_{m_k}\{\vec{X},\vec{Y}\}P_{S_k}\{\vec{X}\oplus\vec{Y}),\vec{Z}\} \\
&\approx \sum_{\vec{Y}\in D(\vec{X},\vec{Z})} C(\vec{X},\vec{Y};\vec{Z})\exp(-V(\vec{X},\vec{Y};\vec{Z})\ln k)
\end{aligned} \tag{3.5.5}$$

其中

$$D(\vec{X},\vec{Z}) = \{\vec{Y};\vec{Z}\subset\vec{X}\oplus\vec{Y}\}$$

$$V(\vec{X},\vec{Y};\vec{Z}) = \sum_{j=1}^{N}[ad(X_j,Y_j) + b(\hat{f}(\vec{X}\oplus\vec{Y}) - f(Z_j))]$$

由于在式(3.5.5) 中，当 $k\to\infty$ 时，$V(\vec{X},\vec{Y};\vec{Z})$ 越小时在式(3.5.5) 中右边的项起作用越大。因此令

$$V(\vec{X},\vec{Z}) = \min_{\vec{Y}\in D(\vec{X},\vec{Z})} V(\vec{X},\vec{Y};\vec{Z})$$

$$\bar{D}(\vec{X},\vec{Z}) = \{\vec{Y};\vec{Y}\in D(\vec{X},\vec{Z}), V(\vec{X},\vec{Y};\vec{Z}) = V(\vec{X},\vec{Z})\} \tag{3.5.6}$$

$$C(\vec{X},\vec{Z}) = \sum_{\vec{Y}\in \bar{D}(\vec{X},\vec{Z})} C(\vec{X},\vec{Y};\vec{Z}) \tag{3.5.7}$$

于是有

$$P\{\vec{X}^k(n+1) = \vec{Z}/\vec{X}^k(n) = \vec{X}\} = C(\vec{X},\vec{Z})k^{-V(\vec{X},\vec{Z})} + o(k^{-V(\vec{X},\vec{Z})}) \tag{3.5.8}$$

由式(3.5.8)可见

$$P\{\vec{X}^k(n+1) = \vec{Z}/\vec{X}^k(n) = \vec{X}\} = 0$$
$$\Leftrightarrow V(\vec{X},\vec{Z}) = \infty \Leftrightarrow C(\vec{X},\vec{Z}) = 0$$

令

$$P\{\vec{X},\vec{Z};k\} = C(\vec{X},\vec{Z})k^{-V(\vec{X},\vec{Z})}$$

由于$|S^N| < \infty$,对于任意$\varepsilon > 0$,存在$K$,当$k \geqslant K$时有

$$(1-\varepsilon)P\{\vec{X},\vec{Z};k\} \leqslant P\{\vec{X}^k(n+1) = \vec{Z}/\vec{X}^k(n) = \vec{X}\} \leqslant (1+\varepsilon)P\{\vec{X},\vec{Z};k\} \tag{3.5.9}$$

**定理 3.5.2** 对于本节所述广义模拟退火遗传算法,若它的参数$N,a,l,b,\delta$满足条件$N > al/b\delta$,则$\{\vec{X}^k(n)\}$渐近概率收敛到全局最优解,即

$$\lim_{k\to\infty}\lim_{n\to\infty} P\{\vec{X}^k(n)\cap M \neq \varnothing\} = 1$$

**证明** 记

$$A = \min\{V(\vec{X},\vec{Y},\vec{Z});\vec{X}\cap M \neq \varnothing,\ \vec{Z}\cap M = \varnothing, \vec{Y}\in D(\vec{X},\vec{Z})\}$$

当$\vec{Z}\cap M = \varnothing$时,$f(Z_j) < f^*\ (j \leqslant N)$,$f^* - f(Z_j) \geqslant \delta$。又$\vec{X}\cap M \neq \varnothing$时,$\bar{f}(\vec{X}\oplus\vec{Y}) = f^*$,于是$\bar{f}(\vec{X}\oplus\vec{Y}) - f(Z_j) \geqslant \delta (j \leqslant N)$。于是必有$A \geqslant bN\delta$,从而$\vec{X}\cap M \neq \varnothing$时有

$$P\{\vec{X}^k(n+1)\cap M = \varnothing/\vec{X}^k(n) = \vec{X}\} \leqslant |S^N| N^N k^{-(bN\delta)} = \alpha_k$$

当 $\vec{X} \cap M = \varnothing$ 时，取 $Z \in M$，$\vec{Z}_0 = (Z,\cdots,Z)$ 以及 $\vec{Y}_0 = (X_1,\cdots,X_{N-1},Z)$，则 $\vec{Z}_0 \cap M \neq \varnothing$，$\vec{Z}_0 \subset \vec{X} \oplus \vec{Y}_0$，于是

$$
\begin{aligned}
& P\{\vec{X}^k(n+1) \cap M = \varnothing / \vec{X}^k(n) = \vec{X}\} \\
& \quad \leqslant 1 - P\{\vec{X}^k(n+1) \cap M = \varnothing / \vec{X}^k(n) = \vec{X}\} \\
& \quad \leqslant 1 - C(\vec{X},\vec{Y}_0,\vec{Z}_0)\exp\{-V(\vec{X},\vec{Y}_0,\vec{Z}_0)\ln k\} \\
& \quad \leqslant 1 - \exp\{-al(\ln k)\} = 1 - k^{-al} = \beta_k
\end{aligned}
$$

显然 $\beta_k < 1(k \geqslant 1)$，且

$$
\begin{aligned}
\lim_{k\to\infty} \frac{\alpha_k}{1-\beta_k} &= \lim_{k\to\infty} \frac{|S^N|N^N k^{-(bN\delta)}}{k^{-al}} \\
&= |S^N|N^N \lim_{k\to\infty} k^{-(bN\delta - al)} = 0
\end{aligned}
$$

由定理3.5.1则证。

**定理3.5.3**　对于本节所述广义模拟退火遗传算法，若它的参数 $a,l,b,\delta$ 满足条件 $a(l-1) < b\delta$，则 $\{\vec{X}^k(n)\}$ 渐近概率强收敛到最优解集，即

$$
\lim_{k\to\infty}\lim_{n\to\infty} P\{[\vec{X}^k(n) \subset M]\} = 1
$$

**证明**　记

$$
\begin{aligned}
B = \min\{ & V(\vec{X},\vec{Y},\vec{Z}); \vec{X} \subset M, \vec{Z} \cap M^c \neq \varnothing, \\
& \vec{Y} \in D(\vec{X},\vec{Z})\}
\end{aligned}
$$

当 $\vec{X} \subset M$，$\vec{Z} \cap M^c \neq \varnothing$，$\vec{Y} \in D(\vec{X},\vec{Z})$ 时必有 $V(\vec{X},\vec{Y},\vec{Z}) \geqslant a + b\delta$，从而 $B \geqslant a + b\delta$，于是 $\vec{X} \cap M^c = \varnothing$ 时有

$$
\begin{aligned}
& P\{\vec{X}^k(n+1) \cap M^c \neq \varnothing / \vec{X}^k(n) = \vec{X}\} \\
& \quad \leqslant |S^N|N^N k^{-(a+b\delta)} = \alpha_k
\end{aligned}
$$

与定理3.5.2类似，取 $Z \in M$ 及 $\vec{Z}_0$ 与 $\vec{Y}_0$，可证 $\vec{X} \cap M^c \neq \varnothing$ 时有

$$
\begin{aligned}
& P\{\vec{X}^k(n+1) \cap M^c \neq \varnothing / \vec{X}^k(n) = \vec{X}\} \\
& \quad \leqslant 1 - C(\vec{X},\vec{Y}_0,\vec{Z}_0)\exp\{-V(\vec{X},\vec{Y}_0,\vec{Z}_0)\ln k\} \\
& \quad \leqslant 1 - k^{-al} = \beta_k
\end{aligned}
$$

显然 $\beta_k' < 1(k \geqslant 1)$，且

$$\lim_{k\to\infty}\frac{\alpha_k}{1-\beta_k'} = |S^N|N^N\lim_{k\to\infty}k^{-(b\delta-a(l-1))} = 0$$

由定理 3.5.1，则证。

定理 3.5.3 的条件过于苛刻，要求个体长度 $l$ 是足够小。为了给出更加宽松的条件，我们采取图论方法。由式(3.5.8) 可见，$V(\vec{X},\vec{Z})$ 在研究收敛中是极端重要的。首先给出几个引理。

**引理 3.5.1**　$V(\vec{X},\vec{Z})$ 具有以下性质：

(1) $\forall \vec{X} \in S^N, \vec{Z} \notin \tilde{\beta}^N$，则 $V(\vec{X},\vec{Z}) > 0$

(2) $\forall \vec{X},\vec{Z} \in S^N, V(\vec{X},\vec{Z}) = 0 \Rightarrow \vec{Z} \subset B(\vec{X})$

(3) $\forall \vec{X} \in \beta^N, \forall \vec{Z} \in S^N; \vec{X} \neq \vec{Z} \Rightarrow V(\vec{X},\vec{Z}) > 0$

其中 $\beta^N$ 为齐次种群，即

$$\tilde{\beta}^N = \{\vec{X} \in S^N; f(X_i) = f(X_j)\}$$

**证明**　(1) $\forall \vec{Y} \in S^N$，当 $\vec{Z} \subset \vec{X} \oplus \vec{Y}, \vec{Z} \notin \tilde{\beta}^N$ 时

$$\tilde{f}(\vec{X} \oplus \vec{Y}) > \frac{1}{N}\sum_{j=1}^{N} f(Z_j)$$

于是 $V(\vec{X},\vec{Y};\vec{Z}) > 0$，从而 $V(\vec{X},\vec{Z}) > 0$。

(2) $V(\vec{X},\vec{Z}) = 0$ 时，存在 $\vec{Y} \in S^N$ 使 $\vec{Z} \subset \vec{X} \oplus \vec{Y}$ 和 $V(\vec{X},\vec{Y};\vec{Z}) = 0$，于是 $\vec{X} = \vec{Y}, f(Z_j) = \tilde{f}(\vec{X} \oplus \vec{Y}) = \tilde{f}(\vec{X})$ $(j \leqslant N)$，从而 $\vec{Z} \subset B(\vec{X})$。

(3) 只须证明，$\forall \vec{Y} \in S^N$ 和 $\vec{Z} \subset \vec{X} \oplus \vec{Y}, V(\vec{X},\vec{Y};\vec{Z}) > 0$。事实上，若 $\vec{X} \neq \vec{Y}$，那么 $V(\vec{X},\vec{Y};\vec{Z}) > 0$。假定 $\vec{Y} = \vec{X} = (i)$，其中 $(i) \in \beta^N$ 是以 $i$ 为个体的齐次种群，$\vec{X} = \vec{Y}$，再由 $\vec{Z} \subset \vec{X} \oplus \vec{Y} = (i)$，则 $\vec{Z} = \vec{X}$，这与 $\vec{Z} \neq \vec{X}$ 矛盾，故只能有 $\vec{Y} \neq \vec{X}$，从而 $V(\vec{X},\vec{Y};\vec{Z}) > 0$。

**定义 3.5.2**　设 $L$ 是一个有限集，其元素用 $i,j,k,m,n$ 表示，$W \subset L$，一个由箭头 $m \to n(m \in L \setminus W, n \in L, n \neq m)$ 组成的图称为 $W$ 图，若它满足：

(1) $\forall m \in L \setminus W$ 是一个且仅仅是一个箭头的始点；

(2) 在图中不存在闭循环，或等价地

(2)′ $\forall m \in L \setminus W$ 存在一列箭头，可引向 $n \in W$。

用 $G(W)$ 表示 $W$ 图的全体，$\forall \vec{X} \in S^N$，$G(\vec{X})$ 表示 $S^N$ 上 $\vec{X}$ 图的全体。对图 $g \in G(\vec{X})$，令

$$\prod{}^k(g) = \prod_{(\vec{Y} \to \vec{Z}) \in g} P_k\{\vec{Y}, \vec{Z}\}$$

$$Q_{\vec{X}}^k = \sum_{g \in G(\vec{X})} \prod{}^k(g) = \sum_{g \in G(\vec{X})} \Big( \prod_{(\vec{Y} \to \vec{Z}) \in g} C(\vec{Y}, \vec{Z}) \Big) \cdot \exp\Big\{ - \sum_{(\vec{Y} \to \vec{Z}) \in g} V(\vec{X}, \vec{Z}) \ln k \Big\} \tag{3.5.10}$$

由式(3.5.9)即得，当 $k \geqslant K_1$ 时有

$$(1-\varepsilon)^{2|S^N|-2} \frac{Q_{\vec{X}}^k}{\sum_{\vec{Y}} Q_{\vec{Y}}^k} \leqslant \mu^k(\vec{X}) \leqslant (1+\varepsilon)^{2|S^N|-2} \frac{Q_{\vec{X}}^k}{\sum_{\vec{Y}} Q_{\vec{Y}}^k} \tag{3.5.11}$$

其中

$$\mu^k(\vec{X}) = \lim_{n \to \infty} P\{\vec{X}^k(n+1) = \vec{X} / \vec{X}^k(n) = Y\}$$

当 $k$ 充分大时，式(3.5.10)中右边求和中起主导作用的量是

$$W(\vec{X}) = \min_{g \in G(\vec{X})} \sum_{(\vec{Y} \to \vec{Z}) \in g} V(\vec{Y}, \vec{Z}) \tag{3.5.12}$$

用 $\hat{G}(\vec{X})$ 表示能达到上述极小值的 $\vec{X}$ 图的全体，则有

$$Q_{\vec{X}}^k = \sum_{g \in \hat{G}(\vec{X})} \Big( \prod_{(\vec{Y} \to \vec{Z}) \in g} C(\vec{Y}, \vec{Z}) k^{-W(\vec{X})} \Big) + o(k^{-W(\vec{X})}) \tag{3.5.13}$$

令

$$D(\vec{X}) = \sum_{g \in \hat{G}(\vec{X})} \prod_{(\vec{Y} \to \vec{Z}) \in g} C(\vec{Y}, \vec{Z}) \tag{3.5.14}$$

$$\underline{W} = \min\{W(\vec{Y}); \vec{Y} \in S^N\}$$

$$W^* = \{\vec{X}; W(\vec{X}) = \underline{W}\}$$

于是 $\forall \varepsilon > 0$,存在 $K_2$,当 $k \geqslant K_2$ 时有

$$(1-\varepsilon)\frac{D(\vec{X})}{\sum\limits_{\vec{Y}\in W^*} D(\vec{Y})} k^{\underline{W}-W(\vec{X})} \leqslant \frac{Q_{\vec{X}}^k}{\sum\limits_{\vec{Y}} Q_{\vec{Y}}^k}$$

$$\leqslant (1+\varepsilon)\frac{D(\vec{X})}{\sum\limits_{\vec{Y}\in W^*} D(\vec{Y})} k^{\underline{W}-W(\vec{X})} \tag{3.5.15}$$

于是,当 $k \geqslant \max(K_1, K_2)$ 时有

$$(1-\varepsilon)^{2|S^N|-1}\frac{D(\vec{X})}{\sum\limits_{\vec{Y}\in W^*} D(\vec{Y})} k^{\underline{W}-W(\vec{X})} \leqslant \mu^k(\vec{X})$$

$$\leqslant (1+\varepsilon)^{2|S^N|-1}\frac{D(\vec{X})}{\sum\limits_{\vec{Y}\in W^*} D(\vec{Y})} k^{\underline{W}-W(\vec{X})}$$

从而得到

$$\lim_{k\to\infty}\mu^k(\vec{X}) = \begin{cases} 0, & \vec{X} \notin W^* \\ \dfrac{D(\vec{X})}{\sum\limits_{\vec{Y}\in W^*} D(\vec{Y})}, & \vec{X} \in W^* \end{cases}$$

**引理 3.5.2** 若 $\vec{X} \notin \beta^N$,则 $\forall X \in B(\vec{X}), V(\vec{X},(X)) = 0$。

**证明** 因为

$$V(\vec{X},(X)) = \min_{\vec{Y}\in D(\vec{X},(X))}\left\{\sum_{j=1}^{N}(ad(X_j, Y_j) + b(\hat{f}(\vec{X}\oplus\vec{Y}) - f(X)))\right\}$$

那么由 $X \in B(\vec{X}) \subset \vec{X}$ 知,$\forall \vec{Y} \in S^N$ 有 $X \in \vec{X} \oplus \vec{Y}$,特别取 $\vec{Y} = \vec{X}$ 即得 $V(\vec{X},(X)) = 0$。

**引理 3.5.3** 关于图有以下结论:

(1) 在 $\hat{G}(\vec{X})$ 中,存在图 $g$,对于 $S^N \setminus \beta^N$ 中的每个种群 $\vec{Y} \neq \vec{X}$,在图 $g$ 中存在箭头 $\vec{Y} \to \vec{Z}$,满足 $V(\vec{X},\vec{Z}) = 0$ 且 $\vec{Z} \in \beta^N$。

(2) 对于 $\vec{X} \in \beta^N$，计算 $W(\vec{X})$ 只须考虑 $g$ 是 $\beta^N$ 上的图。

(3) 若 $\vec{X} \notin \beta^N$，则有

$$W(\vec{X}) = \min_{\vec{Y} \in A}(W(\vec{Y}) + V(\vec{Y},\vec{X}))$$

由引理 3.5.1 和引理 3.5.3(3)，可知 $W^* \subset A$。事实上，由引理 3.5.3 知，当 $\vec{X} \notin \beta^N$ 时

$$W(\vec{X}) \geqslant \min_{\vec{Y} \in A} W(\vec{Y}) + \min_{\vec{Y} \in A} V(\vec{Y},\vec{X})$$

又由引理 3.5.1 知，$\forall \vec{Y} \in \beta^N, \vec{X} \notin \beta^N$ 有 $V(\vec{Y},\vec{Z}) > 0$，故

$$W(\vec{X}) > \min_{\vec{Y} \in A} W(\vec{Y}) \geqslant \min_{\vec{Y} \in S^N} W(\vec{Y})$$

从而有 $\vec{X} \notin W^*$，$W^* \subset \beta^N$。

以下视 $W^*$ 为 $S$ 的子集，并用 $V(X,Y)$，$W(X)$ 分别表示 $V((X),(Y))$ 和 $W((X))$。

**引理 3.5.4**　$a,b$ 是本节所述广义模拟退火算法中的参数，$\forall X,Y \in S$，有

$$V(X,Y) = ad(X,Y) + bN(f(X) - f(Y))^+$$

**证明**　当 $X = Y$ 时等式显然成立。若 $X \neq Y$，取

$$Z = (X,X,\cdots,X,Y)$$

使

$$V((X),(Y),Z) = \sum_{j=1}^{N}[ad(X_j,Z_j) + b(f(\vec{X} \oplus \vec{Z}) - f(Y))]$$

达到极小值 $V(X,Y)$，且当 $f(X) \leqslant f(Y)$ 时，$V(X,Y) = ad(X,Y)$；当 $f(X) > f(Y)$ 时，$V(X,Y) = ad(X,Y) + bN(f(X) - f(Y))$，则证。

**定理 3.5.4**　对于本节所述广义模拟退火遗传算法，当它的参数 $N,a,l,b,\delta$ 满足 $N > a(l-1)/b\delta$ 时，$\forall \vec{X} \in S^N, Z \in M$ 有

$$\lim_{k\to\infty} \lim_{n\to\infty} P\{\vec{X}^k(n+1) = (Z) / \vec{X}(0) = \vec{X}\}$$

$$= \lim_{k\to\infty}\mu^k((Z)) = \frac{1}{|M|}$$

其中

$$\delta = \min\{|f(X) - f(Y)|; f(X) \neq f(Y), X, Y \in S\}$$

**证明** 根据引理 3.5.3(2)，$\forall X \in S$，按式(3.5.12)计算 $W((X))$ 时只须考虑 $\beta^N$ 上的$(X)$图，即 $S$ 上的$X$图。下面分三步证明：

(1) 证明 $W^* \subset M$。

只须证明 $\forall X \in M$ 有

$$W((X)) < \min_{Y \notin M}(W((Y))$$

任取 $S$ 上的 $X$ 图 $g$，由 $X \notin M$ 知，在 $g$ 中存在路径 $X_0 \to X_1 \to \cdots \to X_s \to Y$，其中 $s \geqslant 0, X_i \in M(i \leqslant s), Y \notin M$。我们可以分两步将 $X$ 图 $g$ 变成 $X_0$ 图 $\tilde{\tilde{g}}$。

(a) 将图 $g$ 中 $X_s \to Y$ 删掉，并补加 $X \to X_0$ 得 $X_s$ 图 $\bar{g}$，且由引理 3.5.4 得到

$$\begin{aligned} V(g) - V(\bar{g}) &= ad(X_s, Y) + bN(f(X_s) - f(Y)) \\ &\quad - ad(X, X_0) \\ &> -a(l-1) + bN\delta > 0 \end{aligned}$$

(b) 将 $X_{k-1} \to X_k$ 换成 $X_k \to X_{k-1} (k \in s)$ 得 $X_0$ 图 $\tilde{\tilde{g}}$，由于

$$f(X_0) = \cdots = f(X_s)$$

且 $\forall X, Y \in S, d(X,Y) = d(Y,X)$，故 $V(X_{K_{k-1}}, X_k) = V(X_k, X_{k-1}) (k \leqslant s)$，从而

$$V(g) > V(\bar{g}) = V(\tilde{\tilde{g}}) \tag{3.5.17}$$

又因 $\forall X \in S$，根据引理 3.5.3(2)，计算 $W((X))$ 时只须对 $S$ 上的 $X$ 图$g$ 求极小，故由式(3.5.17)得 $W((X)) > W((X_0))$，从而

$$\min_{X \notin M} W((X)) > W((X_0))$$

于是 $W^* \subset M$。

(2) 证明 $\forall X, Y \in M, W((X)) > W((Y))$ 且 $W^* = M$。

任取 $X$ 图 $g \in \hat{G}((X))$，由 $Y \neq X$，在 $g$ 中存在唯一路径

$$Y = X_0 \to X_1 \to \cdots \to X_s \to X_{s+1} = X \ (s \geqslant 0)$$

先证明 $\forall k \leqslant s, X_k \in M$。若不然，不妨设 $X_k \notin M$，且

$$(X_1, \cdots, X_{k-1}) \subset M$$

在图 $g$ 中删除 $X_{k-1} \to X_k$，补加 $X_{k-1} \to X_1$，得另一个 $X$ 图 $\bar{g}$ 且有

$$\begin{aligned} V(g) - V(\bar{g}) &= V(X_{k-1}, X_k) - V(X_{k-1}, X) \\ &= ad(X_{k-1}, X_k) + bN(f(X_{k-1}) - f(X_k)) \\ &\quad - ad(X_{k-1}, X) \\ &> -a(l-1) + bN\delta > 0 \end{aligned}$$

于是有 $V(\bar{g}) < V(g) = W((X))$，与 $\bar{g} \in G((X))$ 矛盾。故 $\forall k \leqslant s, X_k \in M$。

现在将路径 $Y = X_0 \to X_1 \to \cdots \to X_s \to X_{s+1} = X$ 换成 $X = X_{s+1} \to X_s \to \cdots \to X_1 \to X_0 = Y$，得 $Y$ 图 $\tilde{\tilde{g}}$ 且有 $V(g) = V(\tilde{\tilde{g}})$，从而 $W((X)) \geqslant W((Y))$。由于 $X, Y \in M$ 是任意的，故 $W((X)) = W((Y))$，同时表明上面的 $Y$ 图 $\tilde{\tilde{g}} \in \hat{G}((Y))$ 且 $W^* \subset M$，于是 $W^* = M$。

(3) 证明 $\forall X, Y \in M$，有 $|\hat{G}((X))| = |\hat{G}((Y))|$ 且 $D((X)) = D((Y))$。

由(2)中证明知 $\forall X$ 图 $g \in \hat{G}((X))$，存在唯一的 $Y$ 图 $\tilde{\tilde{g}} \in \hat{G}((Y))$，而且不难看出不同的 $X$ 图 $g \in \hat{G}((X))$ 对应着不同的 $Y$ 图 $\tilde{\tilde{g}} \in \hat{G}((Y))$，反之亦然。这样在 $\hat{G}((X))$ 与 $\hat{G}((Y))$ 之间可通过(2)证明中的方法建立一一对应 $g \leftrightarrow \tilde{\tilde{g}}$，使 $V(g) = V(\tilde{\tilde{g}})$，同时也证明了 $|\hat{G}((X))| = |\hat{G}((Y))|$。

为了证明 $D((X)) = D((Y))$，在 $V(\vec{X}, \vec{Y}; \vec{Z})$ 的表达式中，

取 $\vec{X}=(X_1),\vec{Z}=(X_2),X_1,X_2\in S$,易知只有取 $\vec{Y}=(X_1)$ 时 $V(\vec{X},\vec{Y};\vec{Z})$ 达最小值,即

$$\hat{D}((X_1),(X_2))=(X_1)$$

又因为

$$C((X_1),(X_2))=C((X_1),(X_1);(X_2))$$
$$=\delta_{X_1}(X_2)=\delta_{X_2}(X_1)=C((X_2),(X_1))$$

再利用 $\hat{G}((X))$ 与 $\hat{G}((Y))$ 在(2)中证明的对应关系 $g\leftrightarrow\tilde{\tilde{g}}$ 和式(3.5.14),即得

$$D((X))=\sum_{g\in\hat{G}((X))}\prod_{((X_1)\to(X_2))\in g}c((X_1),(X_2))$$
$$=\sum_{\tilde{\tilde{g}}\in G((X))}\prod_{((X_1)\to(X_2))\in\tilde{\tilde{g}}}c((X_1),(X_2))=D((Y))$$

于是

$$\lim_{k\to\infty}\mu^k((X))=\frac{1}{|W^*|}=\frac{1}{|M|}$$

**推论 3.5.1** 对于本节所述广义模拟退火遗传算法,若它的参数 $N,a,l,b,\delta$ 满足 $N>\frac{a(l-1)}{b\delta}$ 时,则 $\forall\vec{X}\in S^N$ 有

$$\lim_{k\to\infty}\lim_{n\to\infty}P\{\vec{X}^k(n)\subset M\}=1$$

即渐近模拟退火演化算法概率强收敛到全局最优解集。

**证明** 利用定理 3.5.4 直接可证。

## 3-6 遗传算法的停时计算问题

对于遗传算法我们期望知道首次到达最优解集的时间,即停时问题。由于遗传序列是随机的,首次到达的时间也是随机的。这里我们关心的是首次到达时间的期望值。

**定义 3.6.1** 设$\{\vec{X}(n)\}$是遗传算法序列，称

$$T = \min\{n; \vec{X}(n) \cap M \neq \varnothing\}$$

为遗传算法的首达时间。

用 $B(\vec{X}(n))$ 表示 $\vec{X}(n)$ 中达到最大适应值的个体，$b(\vec{X}(n))$ 是 $B(\vec{X}(n))$ 中个体，则

$$T = \min\{n; f(b(\vec{X}(n))) = f^*\}$$

其中

$$f^* = \max\{f(X); X \in S\}$$

如果用 $E(T)$ 表示首达时间 $T$ 的期望值，则

$$P\{T > cT_0\} \leqslant \frac{E(T)}{cT_0}$$

若 $E(T) \leqslant T_0$ 时

$$P\{T > cT_0\} \leqslant \frac{T_0}{cT_0} = \frac{1}{c}$$

独立地使用遗传算法 $k$ 次，则至少有一次在 $cT_0$ 步之前找到最优解的概率不小于 $1 - \frac{1}{c}$。例如 $c = 2, k = 20$，则至少有一次在 $2T_0$ 步之前找到最优解的概率不小于 $1 - 10^{-6}$。

为了估计首达时间期望值，我们给出以下的方法。将 $S^N$ 分解为 $n$ 个互不相交的种群集合 $S_i (i \leqslant n)$，使

$$S^N = \bigcup_{i=1}^{n} S_i, \qquad S_i \cap S_j = \varnothing \ (i \neq j)$$

且对于任意 $\vec{X} \in S_i, \vec{Y} \in S_j$，当 $i < j$ 时，$f(b(\vec{X})) < f(b(\vec{Y}))$。如果遗传算法是杰出者选择，当 $i < j$ 时，$S_j$ 中的种群不可能转到 $S_i$ 中的种群。当把状态空间分解为 $S_1, S_2, \cdots, S_n$ 后，并不能用 $\{S_1, \cdots, S_n\}$ 作为状态空间构成新的马尔可夫链。为此，记

$$p_{ij} = \min\{P(\vec{X}, \vec{Y}); \vec{X} \in S_i, \vec{Y} \in S_j\} \quad (i < j)$$

$P(\vec{X}, \vec{Y})$ 为齐次马尔可夫链$\{\vec{X}(n)\}$的转移概率。同时，令

$$p_{ii} = 1 - \sum_{j=i+1}^{n} p_{ij}$$

则用($p_{ij}$)作为转移矩阵构成群马尔可夫链。如果令

$$q_{ii} = p_{ii} + \sum_{j=i+2}^{n} p_{ij}$$

$$q_{i,i+1} = p_{i,i+1}$$

$$q_{i,i+k} = 0 \qquad (2 \leqslant k \leqslant n - i)$$

$$q_{ij} = 0 \qquad (j < i)$$

则得到修正的群马尔可夫链的齐次转移矩阵。令 $T_{ij}$ 表示从 $S_i$ 到 $S_j$ 的首达时间，则

$$E(T) \leqslant \sum_{i=1}^{n-1} E(T_{i,i+1})$$

而

$$P\{T_{i,i+1} = \tau\} = q_{i,i+1}(1 - q_{i,i+1})^{\tau-1}$$

于是

$$\begin{aligned} E(T_{i,i+1}) &= \sum_{\tau=1}^{\infty} q_{i,i+1}(1 - q_{i,i+1})^{\tau-1} \cdot \tau \\ &= - q_{i,i+1}\left[\sum_{\tau=1}^{\infty}(1 - q_{i,i+1})^{\tau}\right]' \\ &= - q_{i,i+1}\left(\frac{1}{q_{i,i+1}}\right)' = \frac{1}{q_{i,i+1}} \end{aligned}$$

从而

$$E(T) \leqslant \sum_{i=1}^{n-1} \frac{1}{q_{i,i+1}}$$

**例 3.6.1** 计算

$$f(X) = \sum_{i=1}^{l} X_i$$

的最大值，其中 $X_i \in [0,1](1 \leqslant i \leqslant l)$。

显然,最大值是在 $X_i = 1(i \geqslant l)$ 时达到。$f(X)$ 的值只与个体 $X$ 中有多少1而确定。于是我们可以令

$$S_i = \{X; f(X) = i\} \qquad (0 \leqslant i \leqslant l)$$

在 $S_i$ 中有 $C_l^i$ 个个体。于是我们将 $|S^l|$ 中的 $2^l$ 个个体划分为 $l+1$ 个子集合 $S_0, S_1, \cdots, S_l$。

用 $A_{ik}$ 表示 $S_i$ 中个体有 $k$ 个1转为0,而 $i-k$ 个1被保留的事件,$B_{ijk}$ 表示 $S_i$ 中个体的 $l-i$ 个0中有 $k+j-i$ 个0变为1,而 $l-j-k$ 个0保持不变的事件,于是

$$P(A_{ik}) = C_i^k p^k (1-p)^{i-k}$$

$$P(B_{ijk}) = C_{l-i}^{k+j-i} p^{k+j-i} (1-p)^{l-j-k} \quad (0 \leqslant i < j \leqslant l)$$

从而,$i < j$ 时有

$$\begin{aligned} p_{ij} &= \sum_{k=0}^{l} p(A_{ik}) \cdot P(B_{ijk}) \\ &= \sum_{k=0}^{l} C_i^k \cdot p^k (1-p)^{i-k} \cdot C_{l-i}^{k+j-i} p^{k+j-i} (1-p)^{l-j-k} \\ &= \sum_{k=0}^{l} C_i^k C_{l-i}^{k+j-i} p^{2k+j-i} (1-p)^{l+i-j-2k} \\ &= p^{j-i} (1-p)^{l-(j-i)} \sum_{k=0}^{l} C_i^k C_{l-i}^{k+j-i} \left(\frac{p}{1-p}\right)^{2k} \end{aligned}$$

$$p_{ii} = 1 - \sum_{j=i+1}^{l} p_{ij}$$

于是

$$\begin{aligned} q_{i,i+1} &= p(1-p)^{l-1} \sum_{k=0}^{l} C_i^k C_{l-i}^{k+1} \left(\frac{p}{1-p}\right)^{2k} \\ &\geqslant p(1-p)^{l-1}(l-i) \end{aligned}$$

$$\begin{aligned} E(T) &\leqslant \sum_{i=0}^{l-1} \frac{1}{p(1-p)^{l-1}(l-i)} \\ &= \frac{1}{p(1-p)^{l-1}} \sum_{i=1}^{l-1} \frac{1}{i} \end{aligned}$$

$$\leqslant \frac{\lg l + 1}{p(1-p)^{l-1}}$$

令 $p^* = \frac{1}{l}$，则

$$E(T) \leqslant l(\lg l + 1)\left(1 - \frac{1}{l}\right)^{1-l}$$

$$\leqslant l(\lg l + 1)\exp(1)$$

**例 3.6.2** 计算

$$f(X) = \sum_{i=1}^{l} \prod_{j=1}^{i} x_j$$

的极大值，其中 $X_i \in [0,1]$，$(1 \leqslant i \leqslant l)$。

同样，令

$$S_i = \{X \in S; f(X) = i\} \qquad (0 \leqslant i \leqslant l)$$

将 $|S| = 2^l$ 个个体分解为 $l+1$ 个子集合。当 $l = 4$ 时，有

$$S_0 = \{(0000),(0001),(0010),(0011),$$
$$(0100),(0101),(0110),(0111)\}$$

$$S_1 = \{1000),(1001),(1010),(1011)\}$$

$$S_2 = \{(1100),(1101)\}$$

$$S_3 = \{(1110)\}$$

$$S_4 = \{(1111)\}$$

于是 $\quad p_{i,i+1} = p(1-p)^i$

$$E(T) = \sum_{i=0}^{l-1} E(T_{i,i+1}) = \sum_{i=0}^{l-1} \frac{1}{p_{i,i+1}}$$

$$= \frac{1}{p} \sum_{i=0}^{l-1} \left(\frac{1}{1-p}\right)^i$$

$$= \frac{1-p}{p^2}[(1-p)^{-l} - 1]$$

取 $p = c/l(0 < c < l)$，则

$$E(T)=\frac{l^2}{c^2}\left(1-\frac{c}{l}\right)\left[\left(1-\frac{c}{l}\right)^{-l}-1\right]\leqslant l^2\,\frac{e^c-1}{c^2}$$

在以上的例子中，首达时间的期望值被 $l$ 的多项式界定，但这并不总是成立的。

**例 3.6.3**　给出目标函数

$$f(X)=\begin{cases}l+1, & \|X\|_1=l\\ l-\|X\|_1, & \|X\|_1<l\end{cases}$$

仍然令　$S_i=\{X\in S;\ \|X\|_1=i\}\quad(0\leqslant i\leqslant l)$

于是，对于 $i\leqslant l-1$ 有

$$p_{il}=p^{l-i}(1-p)^i$$

$$p_{ii}=1-p_{il}$$

当 $i<l$ 时

$$E(T_{il})=\frac{1}{p_{il}}=\frac{1}{p^{l-i}(1-p)^i}=l^l\left(\frac{1}{l-1}\right)^i\geqslant l^{l-i}$$

$$2^{-l}\sum_{i=0}^{l-1}C_l^iE(T_{il})\geqslant 2^{-l}\sum_{i=0}^{l-1}C_l^il^{l-i}=\left(\frac{l}{2}\right)^l\sum_{i=0}^{l-1}C_l^i\left(\frac{1}{l}\right)^i$$

$$=\left(\frac{l+1}{2}\right)^l-2^{-l}$$

于是首达时间的期望值被指数下界所控制。

# 参考文献

1 Arunkumar S, Chockalingam T. 1993. Genetic search algorithms and their randomized operators. Computers Math Applic, 25: 91～100

2 Aytug H, Koehler G J. 1996. Stopping criteria for finite length genetic algorithms. Informs Journal on computing, 8: 183～191

3 Bhattacharyya S, Koehler G J. 1994. An analysis of non-binary genetic algorithms with cardinality $2^v$ . Complex Systems, 8: 227～256

4 Cerf R. 1998. Asymptotic convergence of genetic algorithms. Adv Appl Prob. 30: 521～550

5 Daris L. 1991. Handbook of genetic algorithms. Van Nostrand Rainhold, NY

6 Davis TE, Principle J. 1991. A simulatied annealing like convergence theory for the simple genetic algorithms, In: Proc 4th Int conf Genetic Algorithms'91. 174～181

7 Davis T E, Principe J. 1993. A Markov chain framework for the simple genetic algorithm. Evolut Comput. 1: 88～269

8 Dawid H. 1994. A Markov chain analysis of genetic algorithms with a state

dependent fitness functions. Complex Systems, 8: 407～417

9 Eiben A E, Aarts E H L, Van Hea K M. 1990. Global convergence of genetic algorithms: A Markov chain analysis. In: Parallel problems solving from nature Berlin: Berlin Press 4～12

10 Francois, O. 1996. Convergence in simulated evolution algorithms, Complex Systems, 10, 311～319

11 Goldberg D E, Deb K. 1991. A compare analysis of section schemes used in genetic algorithms. In: Rawlins G J E, ed. Foundations of genetic algorithms. San Mateo, CA: Morgan Kanfmann, 69～93

12 Gidas B. 1985. Nonstationary Markov chains and convergence of the annealing algorithms. J Statis physics. 39(12): 73～131

13 Goldberg D E, Segrest P. 1987. Finite Markov chain analysis of genetic algorithms, In: Proc Int Conf Genetic Algorithms'87

14 Goldberg D E. 1991. Real-coded genetic algorithms, virtual alphabets, and blocking. Complex Systems. 5: 139～167

15 Goldberg D E, Korb B, Deb K. 1989. Messy genetic algorithms: Motivation analysis, and first results. Complex Systems, 3: 493～530

16 Goldberg D E. 1989. Genetic algorithms-in search, optimization and machine learning. New York: Addiso-Wesley Publishing company, INC

17 Kolonko M. 1995. A generalized crossover operation for genetic algorithms. Complex Systems, 9: 181～191

18 Lawrence D. 1991. Handbook of genetic algorithms. New York: Van Nostrand Reinhold

19 Leung Yee, Mei Chang-Lin, Zhang Wen-Xiu. Statistical tests for spatial non-Stationarity based on geographically weighted regression model. Environment and planning A. (to appear)

20 Leung Yee, Li Guo, Xu Zong-Ben. 1998. A genetic algorithm for the multiple destination routing problems. IEEE Transactions on Evolutionary Computation, 2: 150～161

21 Fischer M, Leung Yee. 1998. A Genetic-Algorithms beased evolutionary computational neural network for modelling spatial interaction data. The

. Annals of Regional Science, 32: 437～458

22 Leung Yee, Gao Yong, Xu Zong-Ben. 1997. Degree of population diversity- A perspective on premature convergence in genetic algorithms and its markov chain analysis. IEEE Transactions on Neural Networks, 8: 1163～1176

23 Leung Yee, Zhang Jiang-She, Xu Zong-Ben. 1997. Global annealing genetic algorithm and its convergence analysis. Science in China (Series E), 40: 414～424

24 Leung Yee, Zhang Jiang-She, Xu Zong-Ben. 1997. Neural networks for convex hull computation. IEEE Transactions on Neural Networks, 8: 601～611

25 Leung Yee, Gao Yong, Xu Zong-Ben. 1996. A markov chain analysis of premature convergence in genetic algorithms. Proceedings of the International Conference on Neural Information Processing, Vol 2. Singapore: Springer-Verlag, 1330～1334

26 Leung Kwong-sak, Leung Yee, So Leo, et al. 1992. Rule learning in expert systems using genetic algorithms: 1, Concepts. Proceedings of Second International Conference on Fuzzy Logic and Neural Network. Vol 1, Japan: 201～204

27 Leung Kwong-Sak, Leung Yee, So Leo, et al. 1992. Rule learning in expert systems using genetic algorithms: 2, Empirical Studies. Proceedings of Second International Conference on Fuzzy Logic and Neural Network. Vol 1, Japan: 205～208

28 Holland J H. 1973. Genetic algorithms and the optimal allocations of trials. SIAM J of Computing, 2: 88～105

29 Holland J H. 1992. Genetic algorithms. Scientific American, 4: 44～50

30 Holland J H. 1995. Adaptation in natural and artficial systems. Michigan: The University of Michigan Press.

31 Horn J. 1993. Finite Markov chain analysis of genetic algorithms with niching. In: Procedings of the fifth international conference on Genetic Algorithm. San Mateo: Morgan Kaufmann, 110～117

32 Naniezzo V. 1994. Genetic evolution of the topology and weight distribution of neural networks. IEEE Trans Neural Network, 5(1): 54～65

33 Mahfound S W. 1993. Finite Markov chain models of an alternative selection strategy for the genetic algorithm. Complex Systems, 7: 155～170

34 Michelewicz Z. 1996. Genetic algorithms + data structure = evolutionary programs, 3rd ed. New York: Springer-Verlag

35 Miller B L, Goldberg D E. 1995. Genetic algorithm, tournament selection, and effects of noise. Complex Systems, 9: 193～212

36 Neveu J. 1975. Discrete-parameter martingales. North-Holland.

37 Nix A E, Vose M D. 1992. Modeling genetic algorithms with Markov chains. Ann math Artificial Intell, 5: 79～88

38 Parsons R J, Forrest S, Burks C. 1995. Genetic algorithms, operators, and DNA fragment assembley. Machine Learning, 21: 11～33

39 Peck C C, Dhawan A P. 1995. Genetic algorithms as global random search methods: An alternative perspective. Evolut Comput, 3(1): 39～80

40 Poli R, Langdon W B. 1998. Schema theory for genetic programming with one-point crossover and point mutation. Evolutionary Computation, 5(3): 231～252

41 Qi X F, Palmieri F. 1994. Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space, I: Basic properties of selection and mutation. IEEE Trans on Neural Networks, 5(1): 96～101

42 Rabinovich Y, Wigderson A. 1991. An analysis of a simple genetic algorithms, In: Proceedings of the Fourth International Conference on Genetic Algorithms. San Mateo: Morgan kamfmann, 251～221

43 Radcliffe N J. 1994. The algebra of genetic algorithms, annals of Maths and artificial Intelligence. 10: 339～384

44 Reguolds D, Gomatam J. 1996. Stochastic modelling of genetic algorithms. Artificial intelligence, 82: 303～330

45 Radcliffe N J. 1991. Equivalence class analysis of genetic algorithms. Complex Systems, 5(2): 181～205

46 Rudolph G. 1994. Convergence properties of canonical genetic algorithms.

IEEE Trans on Neural Networks, 5(1): 96～101

47 Seneta E. 1981. Non-negative matrices and markov chains, 2nd ed, Springer

48 Solomon R. 1997. Raising theoretical guestions about the utility of genetic algorithms. In: Evolutinary programming 4, Lecture Notes in Computer Science, 1213: 275～284

49 Srinivas M, Patnaik L M. 1994. Adaptive probabilities of crossover and mutation in genetic algorithms. IEEE Trans Sys Man and Cybern, 24(4): 656～667

50 Suzuki J. 1995. A Markov chain analysis on simple genetic algoithms. IEEE Trans Sys Man and Cybern, 25: 655～659

51 Suzuki J. 1998. A further result on the Markov chain model of genetic algorithms and its application to a simulated annealing-like strategy. IEEE Trans Sys, Man and Cybern, part B: Cybermeteics: 28(1)

52 Voget S. 1996. Theoretical analysis of genetic algorithms with infinite population size. Complex Systems, 10: 167～183

53 Vose M D, Liepins G E. 1991. Punctuated equilibria in geneitic search. Complex Systems, 5: 31～44

54 Vose M D, Wright A H. 1995. Simple genetic algorithms with linear fitness. Evolutionary computations, 2(4): 347～368

55 Wright A H, Vose M D. 1996. Finiteness of the fixed point set for the simple genetic algorithm, Evolutioary computation

56 Yao X. 1993. Evolutionary artificial neural networks. Int J Neural Systems, 4(3): 203～222

57 Yao X. 1993. A review of evolutionary artificial neural networks. Int J Intelligence Systems, 8: 539～567

58 Zhang J S, Xu Z B, Leung Y. 1996. The genetic algorithms with global annealing and the necessary and sufficient condition of it's convergence. Science in China (Series E) 26(5)

59 Zhigljavsky A A. 1991. Theory of global random search. London: Kluwwer Academic Publishers

60 张文修，梁怡．1996．不确定性推理原理（西安交通大学学术专著丛书）．西安：西安交通大学出版社

61 张文修，汪振鹏，高勇．1996．集值随机过程．北京：科学出版社

62 张文修，梁广锡．1997．模糊控制与系统．西安：西安交通大学出版社

63 高勇．1997．遗传算法的随机模型、理论分析及应用．西安交通大学博士学位论文

64 高勇，张文修．1994．超空间上的选择算子及其在随机过程中的应用．中国科学(A辑)，2：1～10

65 张文修，李腾．1988．集值测度的表示定理．数学学报，2

66 徐宗本，高勇．1996．遗传算法过早收敛现象的特征分析及其预防．中国科学(E辑)，4：364～375

67 张讲社，徐宗本，梁怡．1997．遗传算法的整体退火选择及其收敛主要条件．中国科学(E辑)，2：154～164

68 刘勇，康立山，陈毓屏．1997．非数值并行算法——遗传算法．北京：科学出版社

69 云庆夏，黄光球，王战权．1997．遗传算法和遗传规划．北京：冶金工业出版社

70 Leung Y, Ma J H, Zhang W X. 2001. A new method for mining regression classes in longe data sets. IEEE Trans. Pattern Analysis and Machine Intelligence, 23(1):521

71 Wu W Z, Zhang W X, Wang R M. 2001. Set valued bartle integrals. Journal of Mathematical Analysis and Applications, 255:120

72 Leung Y, Mei C L, Zhang W X. 2000. Testing for spatial autocorrelation among the residuals of the geographically weighted regression. Environment and Planning A, 32:871890

73 Leung Y, Mei C L, Zhang W X. 2000. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environment and planning A, 32:932

74 Leung Y, Gao Y, Xu Z B. 1997. Degree of population diversity – A perspective on premature convergence in genetic algorithm and its Markov chain analysis. IEEE Trans. on Neural Networks, 8(4):11651176

75 Leung Y,Li G,Xu Z B. 1998. A genetic algorithm for multiple destination rounting Problem. IEEE Trans. on Evolutionary Computation,2(4):150-161

76 Leung K S,Duan Q H,Xu Z B, et al. 2001. A new model of simulated evolutionary computation:Convergence analysis and Specifications. IEEE Trans. on Evolutionary Computation,5(1)

77 Gao Y,Xu Z B, Li G. 1999. Theoretical analysis,new algorithms,and some applications of genetic algorithms:A review of some recent work,in "Fuzzy Logic and Soft Computing"(eds. By Chen G Q et al),121135,Kluwer Academic publishers

78 Fan H Y,Lu J W Z,Xu Z B. 2000. An empirical comparison of three novel genetic algorithms. Engineering Computations,17(8):9811001

79 徐宗本,李国. 1995. 解全局优化问题的仿生类算法(1)－模拟进化算法. 运筹学杂志,14(2):111

80 徐宗本,陈志平,章祥荪. 2000. 遗传算法理论研究的新近发展. 数学进展,29(2):97113

81 徐宗本,聂赞坎,张文修. 2001. 关于遗传算法公理化模型的进一步结果. 工程数学学报,18(1): 111